University College London

The Genomic Anchors of the Human Major Histocompatibility Complex

Thesis submitted for the degree of Doctor of Philosophy by

Diego Ottaviani

Cancer Research UK London Research Institute

September 2008

Declaration

I, Diego Ottaviani, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Eukaryotic chromatin is organised into a hierarchy of topologically constrained loop structures. Matrix Attachment Regions (MARs) are genomic sequences that mediate the anchoring of chromatin to the insoluble proteinaceous fraction of the nucleus known as the nuclear matrix. Since only a few MARs have been characterised so far, their role in genomic structure and function is not well understood.

The aim of this thesis is to use the human Major Histocompatibility Complex (MHC) as model region to provide novel insights into the relationship between chromatin folding mediated by MARs and gene expression. This large locus contains critical genes for immunity and is associated with more diseases than any other genomic region. Classical MHC genes are expressed in a cell type specific pattern, and can be induced by cytokines such as IFN-γ.

MARs were identified across the entire MHC in uninduced fibroblasts, IFN-γ induced fibroblasts and B lymphoblastoid cells. Expression array analysis showed that these cell types exhibit different MHC expression profiles. MARs were first isolated treating nuclei with hypertonic buffers followed by nuclease digestion and then mapped by hybridizing them onto a novel tiling path array for the MHC. The suitability of this array platform to study DNA-protein interactions was verified using hybridisations of CIITA-enriched DNA and DNA enriched in H3-K9/K14 acetylation.

The findings reveal that MARs are unevenly distributed across the MHC, and that they can be classified into three classes: constitutive, cell type specific and transcription-dependent. These sequences are mainly positioned in intergenic regions and in close proximity to the MHC class boundaries, subdividing the locus into physical domains. By comparing the position of MARs in uninduced fibroblasts, IFN-γ induced fibroblasts and B lymphoblastoid cells, transcriptional activation of the MHC was found to be associated with a reconfiguration of chromatin architecture resulting from the formation of additional genomic anchors. These results suggest that the dynamic tethering of chromatin is linked to transcriptional regulation.

Acknowledgements

First, I would like to express my deepest gratitude to Denise Sheer. Her scientific vision and contagious enthusiasm have been the source of great inspiration for me. Throughout my PhD, Denise has provided me with sound advice, great teaching, guidance, encouragement and countless inspiring ideas. Denise has also been an exceptional listener and extremely supportive in difficult times and I will forever be indebted to her. I am also grateful to my second supervisor Frank Uhlmann, thesis committee member Ian Tomlinson and collaborators Stephan Beck and Vardhman Rakyan for their helpful advice, to Richard Mitter for being extremely generous with his time with the statistical analysis and Alistair Newall and Rossitza Christova for their guidance. I would like to thank Paul Mulholland for teaching me aCGH and his support and Eleni Tomazou for her essential contribution to the design and construction of the MHC tiling path array and for being a great friend. Many thanks to Cancer Research UK and to all the members of the laboratory, Radost Vatcheva, Pei-Jun Wu, Chiara Mazzanti, Petros Takousis, Badma Segarane, Andi Bolzer, Rossen Donev and William Ogunkolade and also to Cell Services, Equipment Park, the Bioinformatics and Biostatistics group of the London Research Institute and the Microarray Facility of the Sanger Institute. I wish to express my heartfelt gratitude to my very special friend, colleague and younger 'brother' Elliott Lever. I consider myself extremely fortunate to have had the opportunity of working with him. His constructive criticism, advice, kindness and generosity have been invaluable to me. I would also like to thank my friends Tania Jones, Tim Forshew, Ruth Tatevossian, Claire Scott, Tommaso Donati, Lorenzo D'Ercole, Rodolfo Quaresima and flatmates Fran Stressmann and Arthur Talman, for being great listeners, for their entertainment, for keeping me sane and for all their emotional support. I wish to thank my extended family for believing in me and for their unconditional support. In particular, I would like to thank my beloved sister Silvia and grandmother Simonetta for being always there for me, for the care they have provided and for helping me get through difficult times. Lastly, I would like to thank my parents, Maretta and Franco for raising me, supporting me, teaching me, and loving me. Without my parents I could have never followed my dreams. To them I dedicate this thesis.

Table of contents

List of figures

List of tables

Glossary

3C	chromosome conformation capture
aCGH	array comparative genomic hybridisation
APC	antigen presenting cell
ATF	activating transcription factor
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BLS	bare lymphocyte syndrome
BRG1	brahma-related gene 1
BSA	bovine serum albumin
cDNA	complementary DNA
CENPA	centromere protein A
ChIP	chromatin immunoprecipitation
CIITA	class II transactivator
CLIP	class II associated invariant chain peptide
CPI	complete protease inhibitor
CREB	cAMP response element binding protein
cRNA	biotin-labelled antisense RNA
CT	chromosome territory
CTCF	CCCTC-binding factor
CT-IC	chromosome territory-interchromatin compartment
CTL	cytotoxic T-lymphocyte
Cy	cyanine
DAPI	4',6-diamidino-2-phenylindole
dCTP	deoxycytidine triphosphate
dUTP	deoxyuridine triphosphate
DMEM	Dulbecco's modified Eagle's medium
DNMT	DNA methyltransferase
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
ECACC	european collection of cell cultures
EDTA	ethylenediaminetetracetic acid
EM	electron microscope
ER	endoplasmic reticulum

ES	embryonic stem
FCS	foetal calf serum
FISH	fluorescence in situ hybridisation
GATA1	GATA binding protein 1
GEO	gene expression omnibus
GFP	green fluorescent protein
H	histone
H3-K9/14	lysine 9/14 of histone H3
HAT	histone acetyltransferase
HDAC	histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLA	human leukocyte antigen
HMT	histone methyltransferase
hnRNP	heterogeneous nuclear ribonucleoprotein
HRMT	histone arginine methyltransferases
HKMT	histone lysine methyltransferase
HS	hypersensitive site
IC	interchromatin compartment
ICD	interchromosome domain or interchromatin domain
IFN	interferon
Ig	immunoglobulin
IGC	interchromatin granule cluster
Ii	invariant chain
IP	immunoprecipitation
IRF-1/2	interferon response factor 1/2
ISRE	IFN-stimulated response element
JAK	Janus kinase
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LIS	lithium di-iodosalicylic acid
LCR	locus control region
LTA	lymphotoxin-α
LTB	lymphotoxin-β
LTR	long terminal repeat
MAC	membrane attack complex

MAR/SAR	matrix/scaffold attachment region
MeDIP	methylated DNA immunoprecipitation
MHC	major histocompatibility complex
MLS	multi-loop sub-compartment
mRNA	messenger RNA
MSK	mitogen-stress-induced kinases
NCP	nucleosome core particle
NF-kB	nuclear transcription factor kB
NFY	nuclear factor Y
NK	natural killer
NP-40	Nonidet P-40
ORF	open reading frame
ORI	origin of replication
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PML	promyelocytic leukaemia
Pol	polymerase
PSMB	proteasome beta subunit
PRMT1	protein arginine methyltransferase 1
PRP	primer recognition proteins
RNA	ribonucleic acid
RNase	ribonuclease
RT-PCR	real-time PCR
RNP	ribonucleoprotein
RMA	robust multi-chip average
RPMI	Roswell Park Memorial Institute
RW/GL	random-walk/giant-loop
SAFA	scaffold attachment factor A
SAFB1	scaffold attachment factor B1
SATB1	sequence binding protein 1
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
SR	serine/arginine-rich

SSC	saline sodium citrate
STAT	signal transducer and activator of transcription
TE	tris EDTA
TAE	tris acetate EDTA
TFIIC	transcription factor II C
TNF	tumour necrosis factor
TRIM	tripartite motif
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
UCSC	University of California Santa Cruz
UTP	uridine triphosphate
UV	ultraviolet
xMHC	extended MHC

1 Introduction

A newly formed international consortium has recently announced the '1000 Genomes Project', an ambitious effort which involves the sequencing of at least a thousand people from around the world. Although there is no doubt that sequencing projects have been and will continue to be critical for our understanding of genome function, the linear sequence of nucleotide bases does not provide the complete description of the genome as genetic information exists in a dynamic and three-dimensional complex of DNA with proteins. A central question in biology is how the genetic information is packaged into the nucleus so that it can be accessed by nuclear regulatory machineries during cellular differentiation and responses to environmental stimuli. Through hierarchical levels of folding, DNA molecules are packaged with specific proteins into chromosomes, which are compartmentalised into discrete domains called chromosome territories. The degree of chromatin condensation and the attachment of DNA to the nuclear periphery have been shown to contribute to chromosome architecture. Three-dimensional chromosomal structure is also driven by other properties including the affinity of chromatin for nuclear components such as nucleoli and splicing speckles. However, the determinants of chromatin organisation are still largely unknown. The aim of my thesis is to improve our understanding of the relationship between chromatin architecture and gene expression in a model genomic region, the human Major Histocompatibility Complex (MHC).

1.1 The complex architecture of the human genome

A typical nucleus of a human cell is about 5 μm in diameter and contains approximately 2 metres of DNA. This tight packaging is achieved through a complex series of folding levels.

1.1.1 DNA organisation in chromosomes

The DNA of the human genome contains approximately 3.2×10^9 nucleotides distributed over 24 chromosomes (Makalowski, 2001). With the exception of germ-line cells and

specialised cell types lacking DNA, such as red blood cells, each human cell contains 46 chromosomes, of which 22 pairs are autosomes and 1 pair are sex chromosomes.

Each chromosome is characterised by a centromere and two telomeres. The centromere divides the chromosome into the short and long arms designated as p-arm and q-arm, respectively. By interacting with spindle microtubules, the centromere ensures correct segregation during mitosis and meiosis. The telomeres are located at the termini of each chromosome and protect chromosomes from progressive shortening after each mitotic cycle.

The chromosomes of the human genome differ significantly with respect to size. Chromosome 1 is the largest of the human chromosomes and is approximately six times longer than chromosome 21, 22 and Y, which are the three smallest human chromosomes (Gregory et al., 2006). The shape of each chromosome, which depends on the position of its centromere, is also variable.

The structure of human chromosomes undergoes major changes during the cell cycle. At G1-phase, each chromosome is relatively decondensed and is not visible under the light microscope. During S-phase, the DNA molecules are replicated. At G2-phase, each chromosome consists of two sister chromatids attached at the centromere. As the cell progresses into M-phase, the DNA molecules become more highly condensed and form mitotic chromosomes, which are readily visible under the light microscope. Upon completion of mitosis and segregation of the chromatids to each daughter nucleus, each chromosome is progressively decondensed.

By applying specific chemical treatments, each metaphase chromosome can be characterised by a unique band pattern. One of the most common treatments for visualising chromosome bands involves the use of trypsin and the Giemsa dye, which stains bands with different regional genomic features. Giemsa-dark bands (G-bands) are associated with relatively low GC content and gene density, they replicate late in the cell cycle and are often highly condensed during both interphase and metaphase (Niimura and Gojobori, 2002). In contrast, Giemsa-light bands (R-bands) are associated with relative high GC content and gene density, they replicate early in the cell cycle and are less condensed than Giemsa-dark bands (Niimura and Gojobori, 2002). This finding

has revolutionised the identification of chromosomes and the detection of abnormalities that characterise genetic diseases.

1.1.2 Nucleosomal DNA

Within the nucleus, the DNA double helix does not exist as a naked molecule, but rather as a complex with histone (H) and non-histone proteins to form chromatin, which is the substrate of enzymatic machineries involved in critical nuclear processes including DNA replication, transcription and repair. The basic unit of chromatin is the nucleosome, which is formed by DNA wrapped around a disc-shaped structure consisting of highly conserved histone proteins. The individual nucleosome core particle (NCP) consists of an octameric histone structure comprising two copies each of the core nucleosomal histones H2A, H2B, H3 and H4 and 146 nucleotide pairs of double-stranded DNA, which forms 1.7 turns of a left-handed coil (Ramakrishnan, 1997). Nucleosomes are separated by linker DNA of length ranging from a few nucleotide pairs to up to about 80, and can be visualised under the electron microscope as 'beads on a string' structures, which are about 11 nm wide (Khorasanizadeh, 2004; Luger et al., 1997).

The nucleosomal fibre becomes condensed into a 30 nm wide fibre to further compact the DNA. This structure consists of several nucleosomes tightly linked together by the non-nucleosomal H1 histone, a polypeptide characterised by a globular domain that interacts with linker DNA (Ramakrishnan, 1997). The two major models that have been proposed to describe the 30 nm fibre are the solenoid and the zigzag models (Fig 1.1). According to the solenoid model, consecutive nucleosomes are arranged into a simple one-start helix (Finch and Klug, 1976). In the alternative model, the packing of nucleosomes creates a double helical structure (a two-start helix), which twists to produce a zigzag pattern (Bednar et al., 1998).

Higher organisms have genes encoding histone variants that contribute to distinct nucleosomal architectures and confer unique functions to nucleosomes (Brown, 2001). For example the centromere protein A (CENPA), an H3 histone variant, is required for the assembly of centromeric nucleosomes and might function as a centromere mark for kinetochore assembly (Brown, 2001; Howman et al., 2000). While expression of

histone proteins takes place predominantly during S-phase, the synthesis of histone variants occurs throughout the different phases of the cell cycle (Khorasanizadeh, 2004).

Histones and histone variants are characterised by N-terminal and C-terminal tails which extend out from the DNA histone core (Khorasanizadeh, 2004; Ramakrishnan, 1997). These tails are subject to covalent modifications (discussed below), which regulate different aspects of genomic structure and function.

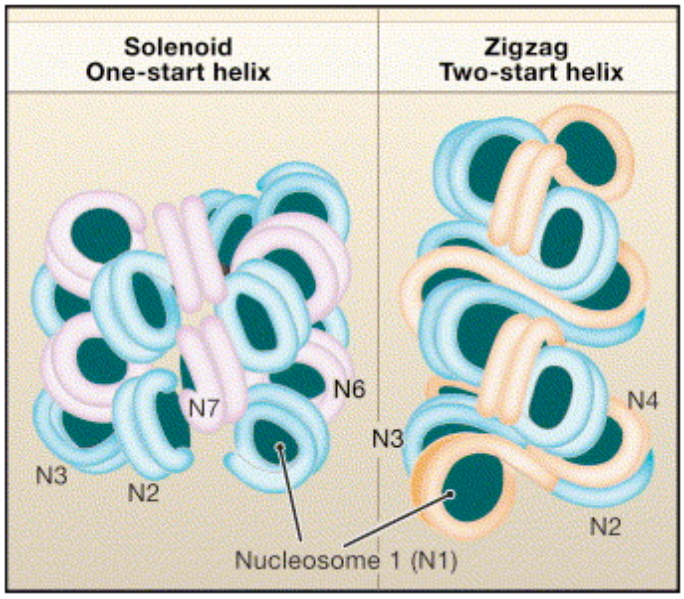

Figure 1.1 Models of the 30 nm chromatin fibre.

(Left) In the solenoid model, the consecutive arrangement of nucleosomes forms an interdigitated one-start helix. Alternative helical gyres are coloured blue and magenta. The linker DNA has not been modelled. (Right) In the zigzag model, the alternating pattern of adjacent nucleosomes generates a two-start helix. Alternative gyres are coloured blue and orange (Tremethick, 2007).

1.1.3 Histone modifications

Histone residues are subject to different covalent modifications. So far, 8 types of modifications have been identified on over 60 histone residues, acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination and proline isomerisation (Kouzarides, 2007). Individual histone modifications and their combinations can change the structure of each nucleosome in many different ways and have the potential to fine-tune mechanisms regulating genomic function.

Histone modifications are catalysed by chromatin remodelling enzymes and can both alter chromatin architecture and recruit specific proteins (Kurdistani and Grunstein, 2003). Genome-wide studies have suggested that these modifications are not evenly distributed across the genome (Kurdistani et al., 2004; Quina et al., 2006). Furthermore, they can dynamically appear and disappear on the nucleosomes depending on the presence or absence of environmental stimuli at the cell surface (Li et al., 2007; Quina et al., 2006). The role of the different modifications is being unveiled by examining their associations with different nuclear processes (Kouzarides, 2007). The most common histone modifications, acetylation, methylation and phosphorylation, which are also the most well characterised, are discussed below. The target residues of each modification type and their relative functions are summarised in Table 1.1.

Histone acetylation is believed to facilitate transcription since it is frequently associated with promoters and coding regions of transcriptionally active genes (Grunstein, 1997; Kurdistani et al., 2004). Acetylation of specific residues on newly synthesised histones at S-phase persists throughout G2-phase if the DNA is damaged, but disappears if the DNA is undamaged, suggesting that this modification could be involved in DNA repair (Celic et al., 2006; Maas et al., 2006). There is also evidence showing that histone acetylation is required for S-phase initiation and chromatin decondensation (Aggarwal and Calvi, 2004; Iizuka et al., 2006; Shogren-Knaak et al., 2006).

Histone methylation can either activate or repress transcription depending on which histone residue is modified. Specific methylated residues at the 5' end and 3' ends of active genes recruit the initiating and the elongating form of RNA polymerase (Pol) II, respectively, and correlate with transcriptional activation (Kouzarides, 2007).

Conversely, other methylated residues are associated with the formation of silent heterochromatin and the repression of euchromatic genes (Kouzarides, 2007). Recent evidence suggests that histone methylation may recruit the cell cycle protein Crb2 to sites of DNA damage and trigger the arrest of the cell cycle at G2-/M-phase to allow DNA repair (Botuyan et al., 2006).

Histone phosphorylation has been linked to transcriptional activation of specific genes, including *FOS* and *JUN* (Clayton et al., 2000; Zippo et al., 2007). In addition, histone phosphorylation contributes to the recognition of sites where DNA repair needs to take place, by marking long stretches of nucleotides around sites of DNA damage (Fillingham et al., 2006). Phosphorylation of specific histone residues facilitates the removal of H1, indicating that this histone modification might also play a role in the accessibility of chromatin (Fischle et al., 2005).

Chromatin Modification	**Residue(s) Modified**	**Functions Regulated**
Acetylation	Lysine	Transcription, Repair, Replication, Condensation
Methylation	Lysine, Arginine	Transcription, Repair
Phosphorylation	Serine, Threonine	Transcription, Repair, Condensation
Ubiquitylation	Lysine	Transcription, Repair
Sumoylation	Lysine	Transcription
ADP ribosylation	Glutamic acid	Transcription
Deimination	Arginine	Transcription
Proline Isomerisation	Proline	Transcription

Table 1.1 Different types of histone modifications.

1.1.4 Histone-modifying enzymes

Histone acetylation is catalysed by histone acetyltransferases (HATs), which transfer an acetyl group from Acetyl CoA to histone lysine residues. HATs are grouped into three main families: GNATs, p300/CBP and MYST depending on their histone lysine specificities (Lee and Workman, 2007). HATs frequently exist in multiprotein complexes and are able to regulate other histone modifications (Baker and Grant, 2007; Lee and Workman, 2007). This can be seen from acetylation of lysine residues, preventing their modification by other histone modifying enzymes and *vice versa* (Yamada et al., 2005; Yang and Seto, 2007). HATs can also affect modifications of neighboring residues (Ahn et al., 2006; Li et al., 2006).

Histone methyltransferases (HMTs) catalyse histone methylation by transferring one to three methyl groups from the cofactor S-adenosyl methionine to lysine or arginine residues. They are divided into histone arginine methyltransferases (HRMTs) and histone lysine methyltransferases (HKMTs) (Berger and Gaudin, 2003; Kouzarides, 2007). Specific HMTs can interact with HATs to synergistically augment transcription as well as recruit DNA methyltransferase (DNMT) to promote DNA methylation (Zhang and Reinberg, 2001). One of the best-characterised HMTs is arginine methyltransferase 1 (PRMT1), which is associated with transcriptional activation (Lin et al., 1996; Tang et al., 2000; Wang et al., 2001).

Studies with specific MAP kinase inhibitors show that histone phosphorylation is mediated by distinct signalling pathways such as the ERK or p38 MAP kinase pathways (Kouzarides, 2007; Soloaga et al., 2003). Downstream kinases, including the mitogen-stress-induced kinases (MSKs), but not the MAP kinases themselves, catalyse the addition of a phosphate group to histone residues (Soloaga et al., 2003). By modifying histones, MSKs promote transcription and regulate spindle function during mitosis and meiosis. The mechanism with which histone modifying kinases catalyse other modifications remains to be established (Nowak and Corces, 2004).

The paucity of isolated and characterised enzymes catalysing histone modifications has limited our understanding of how they regulate chromatin function and structure. (Kouzarides, 2007). While there is evidence linking these enzymes with transcription,

replication and DNA repair processes, the molecular mechanisms underlying such associations remain unclear.

Most covalent modifications of histone tails are reversible and a number of enzymes removing those modifications have been identified (Cheung et al., 2000; Kouzarides, 2007). Histone deacetylases (HDACs), which can catalyse the cleavage of acetyl groups from histone residues, are the best-characterised (Yang and Seto, 2007). Histone modifications are thus governed by the interaction between enzymes adding and removing functional groups from histone residues, enabling cells to alter the structural and functional properties of the genome in response to changing environmental signals.

1.1.5 Heterochromatin and euchromatin

Early microscopic observations revealed two types of chromatin, heterochromatin and euchromatin. Heterochromatin was defined as chromatin remaining highly condensed as the cell makes the transition from metaphase to interphase, and euchromatin as chromatin with a more 'open' conformation. Each of these two chromatin states is characterised by distinctive biochemical features, most importantly histone modifications, which alter the level of chromatin packaging and convert heterochromatin into euchromatin and *vice versa* (Jacobs et al., 2001; Meneghini et al., 2003; Richards and Elgin, 2002; Schneider and Grosschedl, 2007).

Heterochromatin contains transcriptionally silent DNA and euchromatin contains actively transcribing DNA sequences (Schneider and Grosschedl, 2007). While heterochromatin is associated with the NCPs, euchromatin is characterised by nucleosome-free DNase hypersensitive sites (HSs) (Grewal and Elgin, 2002; Sun et al., 2001). Typically, if an active gene from a euchromatic region is relocated to a heterochromatic domain via chromosome rearrangement or transposition, it is usually silenced (Richards and Elgin, 2002). These observations gave rise to the hypothesis that highly packaged chromatin would prevent access of DNA to regulatory factors and would thus function as a physical barrier for transcriptional activation (Dillon and Festenstein, 2002; Misteli, 2007).

The view of chromatin folding as a regulator of accessibility has been challenged by Gilbert et al. (2004) on the basis that chromatin fibres with a more 'open' structure were associated with gene-rich regions while more decondensed chromatin fibres were associated with gene-poor regions (Gilbert et al., 2004). Surprisingly, their analysis also indicated that an 'open' chromatin conformation did not correlate with high levels of gene expression and that a 'closed' conformation could contain active genes. Taken together, these findings suggest a relationship between chromatin architecture and gene density rather than between chromatin architecture and gene activity. Furthermore, Gilbert's findings were in agreement with studies showing that dextrans and proteins with molecular weights of several hundred kilodaltons were able to freely diffuse throughout heterochromatic as well as euchromatic domains within the nucleus (Cremer and Cremer, 2001; Lukacs et al., 2000; Verschure et al., 2003). From these observations, it can be concluded that chromatin condensation does not constitute an insurmountable barrier for transcriptional activation (Misteli, 2007).

1.1.6 Models of higher-order chromatin structure

The higher-order three-dimensional architecture of chromatin structure, which includes the different levels of DNA organisation beyond the 30 nm chromatin fibre up to metaphase chromosomes, is still poorly described. Although numerous models describing higher-order chromatin architecture in both interphase and metaphase nuclei have been put forward, there is still no conclusive evidence for any of them.

Two key models of interphase higher-order chromatin structure are the random-walk/giant-loop (RW/GL) and the multi-loop sub-compartment (MLS) models (Fig 1.2) (Munkel et al., 1999; Sachs et al., 1995). Both of these models interpreted data obtained by employing fluorescence *in vitro* hybridisation (FISH) to visualise distinct chromatin sites and analyse their physical distance as a function of genomic distance. In the RW/GL model, chromatin loops averaging about 3 Mb are associated at their bases to a protein scaffold that extends throughout the nuclear interior. According to this model, the giant loops and the protein scaffold are described as random walks. In the MLS model, the 30 nm chromatin fibre is folded into loop domains of approximately 100 kb. About ten of these loops are spatially arranged in a rosette-like structure forming a sub-compartment of approximately 1 Mb. At the centre of each rosette, the bases of the

loops are held together by a 'loop base spring', which simulates the anchoring function of the nuclear scaffold postulated by the RW/GL model. Multiple consecutive rosettes are interconnected by stretches of 30 nm chromatin fibre and are backfolded to form interphase chromosomes. The MLS model is compatible with studies on DNA replication showing that 1 Mb chromatin domains are visible throughout subsequent cell cycles.

A third model, the chromonema model, was proposed by Belmont et al. (1994) following a study using both light and electron microscopy (EM). Here, the 30 nm chromatin fibre is folded onto itself to form supercoiled, highly condensed structures called chromonema fibres (Fig 1.2) (Belmont and Bruce, 1994). The observed diameters of chromonema fibres are predominantly 60-80 nm in late G1- and early S-phase and 100-130 nm in early G1-phase. Interphase nuclei also have local segments of more highly compacted chromatin measuring 200-400 nm in diameter, as well as more local decondensed chromatin corresponding to more loosely folded chromonema fibres and extended 30 nm chromatin fibres. The folding of the 30 nm chromatin fibre is proposed to be mediated by proteins of a nuclear scaffolding or non-specific inter-fibre interactions at multiple sites and not necessarily at the bases of the loops.

An earlier model, the radial loop model was proposed for mitotic chromosome architecture based on the visualisation of EM micrographs of thin sections containing swollen metaphase chromosomes isolated using different ionic conditions (Marsden and Laemmli, 1979; Paulson and Laemmli, 1977). Here, highly condensed chromatin loops are arranged in a radial fashion with their bases anchored to a central axis extending along most of the length of the chromatid. Recent studies have indicated that the metaphase chromosome axis is constituted of structural proteins which are physically disconnected from one another to form a discontinuous network (Kireeva et al., 2004; Marko, 2008). However, the validity of the radial loop model has been questioned by concerns that the structure of mitotic chromosomes visualised using non-physiological ionic conditions might not reflect their true structure *in vivo*.

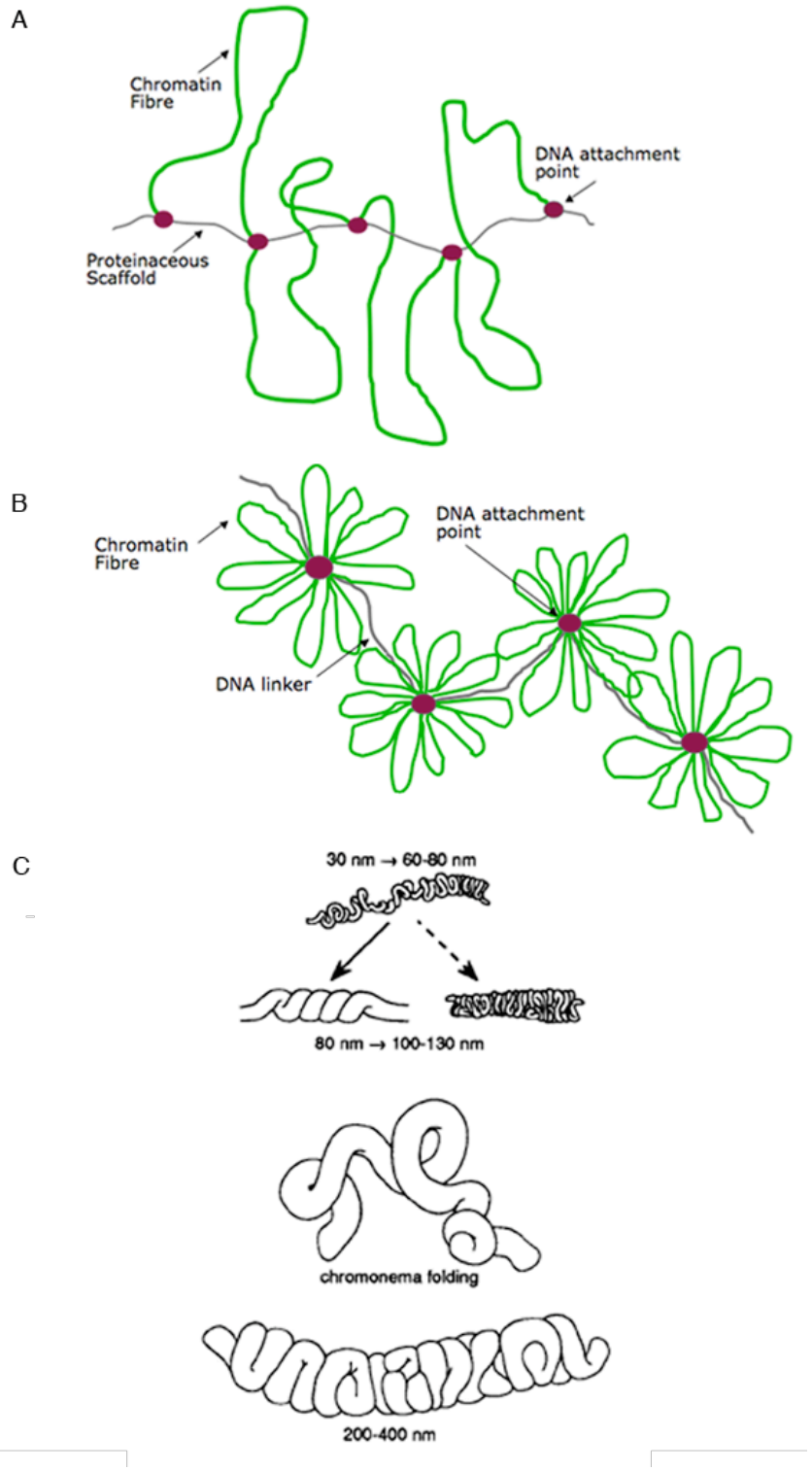

Figure 1.2 Models of higher-order chromatin architecture in interphase cells.

A) The random-walk/giant-loop model proposes that 3 Mb loops are attached to a non-DNA scaffold. **B)** The multiloop sub-compartment model proposes a structure of multiple rosettes containing loops of approximately 100 kb each. **C)** The chromonema model proposes hierarchical levels of chromatin folding back on itself to form fibres with diameters of 60-80 and 100-130 nm, depending on the compaction.

1.1.7 Organisation of chromosomes in the nucleus

It was long assumed that, when cells are not dividing, chromatin exists as long thin tangled threads randomly arranged within the nucleus. Now there is solid evidence that interphase chromosomes display a high degree of organisation (Bolzer et al., 2005; Branco and Pombo, 2007; Cremer and Cremer, 2006; Cremer et al., 2004; Spector, 2003). Groundbreaking experiments by Cremer et al. (1982) using UV microirradiation on defined areas of interphase nuclei in living cells revealed that only a small number of chromosomes were damaged. This indicated that the genome is compartmentalised during interphase (Cremer et al., 1982). The subsequent development of FISH using probes specific for entire chromosomes allowed the visualisation of interphase chromosomes *in situ* and demonstrated that each chromosome occupies a distinct physical domain within the nucleus called a chromosome territory (CT) (Fig 1.3) (Bolzer et al., 2005; Cremer and Cremer, 2006).

Microscopic analysis has provided evidence that CTs are not solid structures but rather are porous entities with invaginations of non-chromatin spaces extending from their peripheries throughout their interiors (Cremer and Cremer, 2001; Misteli, 2005; Verschure et al., 1999; Visser et al., 2000). These spaces might facilitate access of regulatory protein complexes to DNA sequences located within the interior of the CT. Visualisation of chromatin with H2B tagged with green fluorescent protein (GFP) has enabled the identification of interchromatin channels, now called the interchromatin compartment (IC) or interchromatin domain (ICD), starting from nuclear pores, expanding between CTs and into their interior (Cremer and Cremer, 2001; Kanda et al., 1998).

The positioning of CTs is cell type specific. In several human cell types including lymphocytes and lymphoblastoid cell lines, gene-rich chromosomes are situated in the nuclear interior and gene-poor chromosomes in the periphery of the nucleus (Boyle et al., 2001; Cremer et al., 2001; Croft et al., 1999). Gene-dense regions are typically transcriptionally active and replicate early. However, it is still not clear whether in these cell types CT position is shaped by either gene density *per se*, replication timing, transcriptional activity or different combinations of these related parameters (Cremer and Cremer, 2006). In contrast, chromosome size appears to be the major determinant of CT positioning in human fibroblasts since, independent of gene density, larger

chromosomes are preferentially located towards the nuclear periphery and smaller chromosomes towards the nuclear interior (Bolzer et al., 2005).

Branco and Pombo (2006) developed a novel FISH procedure (cryo-FISH), which provides high-resolution images of chromosomes while preserving chromosomal architecture (Branco and Pombo, 2006). The findings revealed that CTs are not entirely physically separated by the IC as they show extensive levels of intermingling mostly at their boundaries. From these data, it was estimated that approximately 20% of the nuclear volume is occupied by intermingling chromosomes. Furthermore, it was shown that in human lymphocytes, the degree of intermingling between specific chromosome pairs is positively correlated with the frequency of chromosomal translocations in the same cell type (Branco and Pombo, 2006; Branco and Pombo, 2007). CT intermingling might thus facilitate interchromosomal interactions and have a direct impact on the formation of chromosomal aberrations.

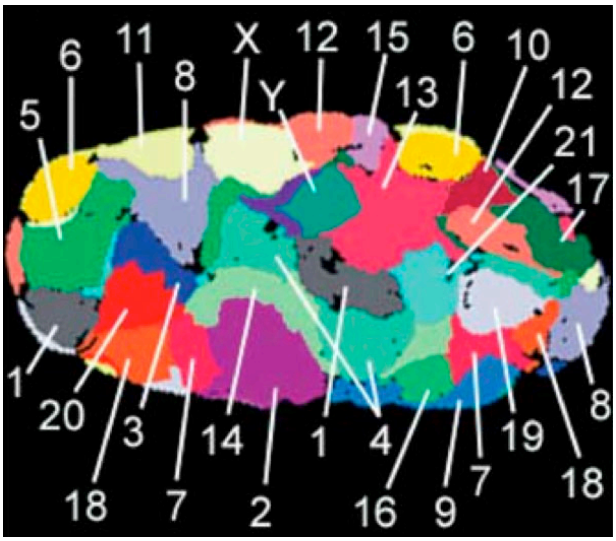

Figure 1.3 Organisation of chromosomes in a human fibroblast nucleus at interphase.

The chromosomes occupy discrete domains called chromosomal territories, which are represented here by different colours (Bolzer et al., 2005).

1.1.8 The nuclear position of genes

The architecture of the nucleus is highly complex since it comprises a nuclear envelope, distinct higher-order chromatin domains such as heterochromatin and euchromatin, and various proteinaceous structures including the nucleolus and small nuclear bodies. As the position of genes relative to these different nuclear landmarks as well as to each other is non-random, it is essential to understand how the spatial organisation of genes affects their function (Misteli, 2005; Parada et al., 2004; Wang et al., 2004).

One of the difficulties in designing experiments to understand nuclear organisation is that there are currently few options for defining a nuclear address. One approach is to determine the radial distance of the gene from the geometrical centre of the nucleus (Lukasova et al., 1997). Another method is to determine the position of the gene within one of the concentric 'shells' of the nucleus of equal radius or volume (Murmann et al., 2005). A novel approach was taken by Wang et al. (2004) who determined the proximity of genes to promyelocytic leukaemia (PML) nuclear bodies (Wang et al., 2004).

The nuclear periphery contains significant amounts of heterochromatin, and therefore it is traditionally regarded as a transcriptionally repressive environment (Misteli, 2004). The repositioning of the murine *IgH* and *Igκ* loci from the nuclear periphery to the nuclear interior following their activation during lymphocyte development supports this view (Kosak et al., 2002). However, recent observations suggest that the role of the nuclear periphery is more complex than originally thought. Experimental positioning of human chromosomes to the nuclear periphery by tethering them to a protein of the inner nuclear membrane was shown to have a silencing effect on many, but not all active genes (Finlan et al., 2008). Furthermore, in budding yeast boundary activities that protect the expression status of active genes, are achieved by binding of active chromatin to nuclear pore complexes (NPCs) (Ishii et al., 2002). Thus, the nuclear periphery is not simply a transcriptionally repressive zone but a heterogeneous nuclear microenvironment regulating gene activation and repression.

The proximity of loci to particular nuclear structures is also of functional significance. Actively transcribed genes frequently colocalise with nuclear structures called interchromatin granule clusters (IGCs), which contain splicing components such as

small nuclear ribonucleoprotein particle (snRNP) and SR proteins (Lamond and Spector, 2003). This has led to the idea that the localisation of active genes near ICGs facilitates transcription by providing relatively high concentrations of splicing components (Kosak and Groudine, 2004). Similarly, the frequent association of a subset of active genes with Cajal bodies, which contain snRNPs, has suggested that this colocalisation is linked to mRNA processing (Gall, 2000; Ogg and Lamond, 2002).

The formation of spatial gene clusters indicates that certain gene loci are non-randomly positioned relative to each other. In a typical mammalian nucleus, 50-100 ribosomal genes located on 2-4 different chromosomes coalesce in the nucleolus where they are brought into physical proximity (Olson et al., 2000). A similar phenomenon was observed for tRNA genes in *Saccharomyces cerevisiae* where over 50 genes from 16 chromosomes are clustered near the nucleolus (Thompson et al., 2003). Studies in erythroid cells showed that certain co-ordinately regulated genes are relocated at shared sites of ongoing transcription upon activation (Osborne et al., 2004). Although these associations could have a regulatory function, it remains to be established whether they are a general feature for other co-regulated genes. The extent to which the physical proximity of particular genes facilitates co-ordinate regulation remains to be established.

In addition to playing a role in gene expression, non-random gene positioning may be involved in the formation of chromosomal abnormalities such as chromosomal translocations (Misteli, 2005). There are strong indications that the positions of genes relative to each other determine their probability of undergoing translocations. The loci involved in translocations in promyelocytic leukaemia, acute myelocytic leukaemia, Burkitt's lymphoma and thyroid lymphoma are observed in closer spatial proximities than non-translocating regions in the corresponding normal cells (Misteli, 2004; Neves et al., 1999; Parada et al., 2004; Roix et al., 2003). These findings are also supported by studies in human lymphocytes showing that the degree of chromosomal intermingling between certain chromosomes correlates with the frequencies of chromosomal translocations (Branco and Pombo, 2006).

Studies carried out so far have investigated the nuclear positions of relatively small genomic regions in a limited number of cell types. The use of high-throughput imaging methods and pattern recognition tools are likely to overcome this limitation and extend

our understanding of the relationship between gene positioning and function (Misteli, 2004).

1.1.9 Chromatin loop structures

Higher-order chromatin architecture is derived from hierarchical loop configurations, which are thought to be of fundamental importance for the regulation of gene expression. Within a eukaryotic cell, the size of these chromatin loops is reported to vary extensively. Local chromatin loops arise from short-range chromatin interactions and can be analysed using biochemical approaches, such as the chromosome conformation capture (3C). Giant chromatin loops can be identified by FISH.

In order to understand the function of chromatin loop domains, it is critical to know how they are formed and determine whether they are either static or dynamic. The development of the 3C technique has provided significant insights into the role of local chromatin loops. This assay involves formaldehyde fixation of cells followed by restriction enzyme digestion, ligation and polymerase chain reaction (PCR) analysis to quantify the frequency of ligated DNA products (Cullen et al., 1993; Dekker et al., 2002). As two restriction fragments in close physical proximity will have a higher ligation frequency than two other restriction fragments located further apart from one another, the *in vivo* 3-D organisation of chromatin can be examined.

These studies have identified interactions of locus control regions (LCRs) or enhancers with their relative promoters and genes to promote gene expression. A classical example is found at the murine β-globin locus where the promoter and the LCR, which is located several kb upstream, are brought into physical proximity by a local rearrangement of the chromatin at the onset of transcriptional activation (Apostolou and Thanos, 2008; Vakoc et al., 2005). The tethering of the loop at its base requires the recruitment of haematopoietic transcription factor GATA1 and its co-factor FOG1, indicating that expression of the locus is associated with loop formation (Vakoc et al., 2005).

Local loops are also formed by the transcription-dependent interaction between promoter and terminator regions of individual genes. Activation of several loci in

Saccharomyces cerevisiae and in humans, is dependent on the folding back of genes onto themselves to juxtapose their 5' and 3' end sequences (Ansari and Hampsey, 2005; O'Sullivan et al., 2004). Physical interactions have also been observed between the 3' end processing complexes and transcription initiation factors (Bentley, 2005; Misteli, 2007). These loops facilitate a single RNA Pol II molecule re-initiating transcription, which would not be possible with a linear configuration since *de novo* recruitment of RNA Pol II molecules would be required for consecutive rounds of transcription (Ansari and Hampsey, 2005).

The physical interaction between insulator DNA sequences which shield genes from stimulatory and repressive effects arising from flanking regions, can also produce local loop structures. In *Drosophila melanogaster*, for instance, the insulators *scs* and *scs'* which are located about 15 kb apart from each other are brought into close proximity by the interaction between *scs* and *scs'* insulator proteins, Zw5 and BEAF, respectively (Blanton et al., 2003). The topological constraints of the loops are thought to establish chromatin domains characterised by unique properties. In *Schizosaccharomyces pombe*, the tethering of insulators with the transcription factor TFIIC at the nuclear periphery mediates the formation of loops containing active sequences which protrude into the nuclear interior (Noma et al., 2006). Here, the loops appear to function as physical barriers separating active euchromatic domains from repressive heterochromatic regions.

Giant chromatin loops have been observed at certain genomic regions with high gene density. This phenomenon was first discovered for the 4 Mb major histocompatibility complex (MHC), which was shown to escape from the borders of the FISH-painted chromosome 6 domain by forming a giant chromatin loop extruding typically about 1 or 2 μm into the surrounding nucleoplasm (Volpi et al., 2000). Upon transcriptional up-regulation of the MHC, the frequency with which the MHC was located on a giant loop increased (Christova et al., 2007; Volpi et al., 2000). Similar higher-order chromatin configurations were subsequently observed in co-ordinately regulated and other gene-rich regions of the genome, such as the human epidermal differentiation complex and the mouse Hox gene clusters (Chambeyron and Bickmore, 2004; Sproul et al., 2005; Williams et al., 2002). Although the molecular basis underlying the formation of these loops and their functional significance are still not clear, it was recently found that the MHC loops represent regions of decondensed chromatin conformation which might

facilitate the binding of transcription factors and promote up-regulation of gene expression (Christova et al., 2007).

1.2 Matrix attachment regions

It has long been hypothesised that the major determinants of chromosomal architecture are found in the fraction of the nucleus termed the nuclear matrix. These determinants are thought to interact with genomic sequences called matrix attachment regions (MARs), located at the bases of chromatin loops, thus organising the genome into loop domains. Although the nuclear matrix is almost certainly not a discrete entity, there is increasing evidence showing that its components are required to establish the three-dimensional structure of chromosomes, organise the nucleus into different functional compartments and regulate fundamental nuclear processes including transcription and DNA replication.

1.2.1 The nuclear matrix

The nuclear matrix is the insoluble fraction that remains after the nucleus is treated with high-strength ionic buffers and nucleases to remove soluble proteins and chromatin. It was first observed approximately 40 years ago where it was described as a residual nuclear protein fraction (Wang, 1966). Although the terms 'nuclear scaffold' and 'nucleoskeleton' are sometimes employed to describe this fraction, the term that is most widely used is nuclear matrix, which was coined by Berezney and Coffey in 1974 (Berezney and Coffey, 1974).

Electron micrographs of the nuclear matrix indicate that it has a fibrogranular organisation extending from the nuclear envelope throughout the nuclear volume (Fig 1.4) (Jackson, 2003; Nickerson, 2001). This structure has been identified in different eukaryotes ranging from yeasts to mammals. It is composed of hundreds of different proteins and RNA and can be divided into two main regions: the nuclear lamina and the internal nuclear matrix that is connected to the lamina and fills the nuclear interior (Berezney and Coffey, 1977). The internal nuclear matrix consists of an intricate ribonucleoprotein (RNP) network of highly branched 10 nm filaments, which are preserved through the matrix isolation procedure (Jackson, 2003). Structural remnants of nucleoli can be observed in this network.

The most common experimental procedures for isolating nuclear matrices are extraction with buffers containing 25 mM lithium di-iodosalicylic acid (LIS), 2 M NaCl or 0.65 M $(NH_2)_2SO_4$ (Donev et al., 2002). The LIS and ammonium sulphate extraction conditions are relatively mild and more physiological compared to the 2 M NaCl procedure (Donev, 2000). However, these and other similar procedures generate underlying networks of core filaments with remarkably similar ultrastructural features (Nickerson, 2001).

Upon treatment with high-salt buffers, soluble proteins including histones are released from the nucleus causing the DNA to unfold and loop out far beyond the nuclear lamina (Vogelstein et al., 1980). These preparations can be observed under fluorescent light microscopy by staining with 4',6-diamidino-2-phenylindole (DAPI), which fluoresces brightly when it forms complexes with the minor groove of double-stranded DNA. Multiple DNA loops, which can be seen as circular 'halos' around the nuclei, remain attached to the insoluble intranuclear structure suggesting that the core filaments of the nuclear matrix are bound to chromatin (Heng et al., 2004; Jackson, 2003).

Whether the nuclear matrix is a continuous network of fibres is still under debate. GFP tagging of *Drosophila* nuclear matrix proteins CP60 and CP190 shows that they form filaments throughout the nucleus *in vivo* (Oegema et al., 1997). In mammalian cells, immunostaining of proteins that resist high-salt extraction has allowed the visualisation of networks of filaments at the nuclear periphery as well as 'cage-like' structures throughout the nucleus (Cai et al., 2003; Dechat et al., 2008). However, most of the characterised nuclear matrix proteins have not been demonstrated to form core filaments in living cells (Jackson et al., 2003; Pederson, 2000).

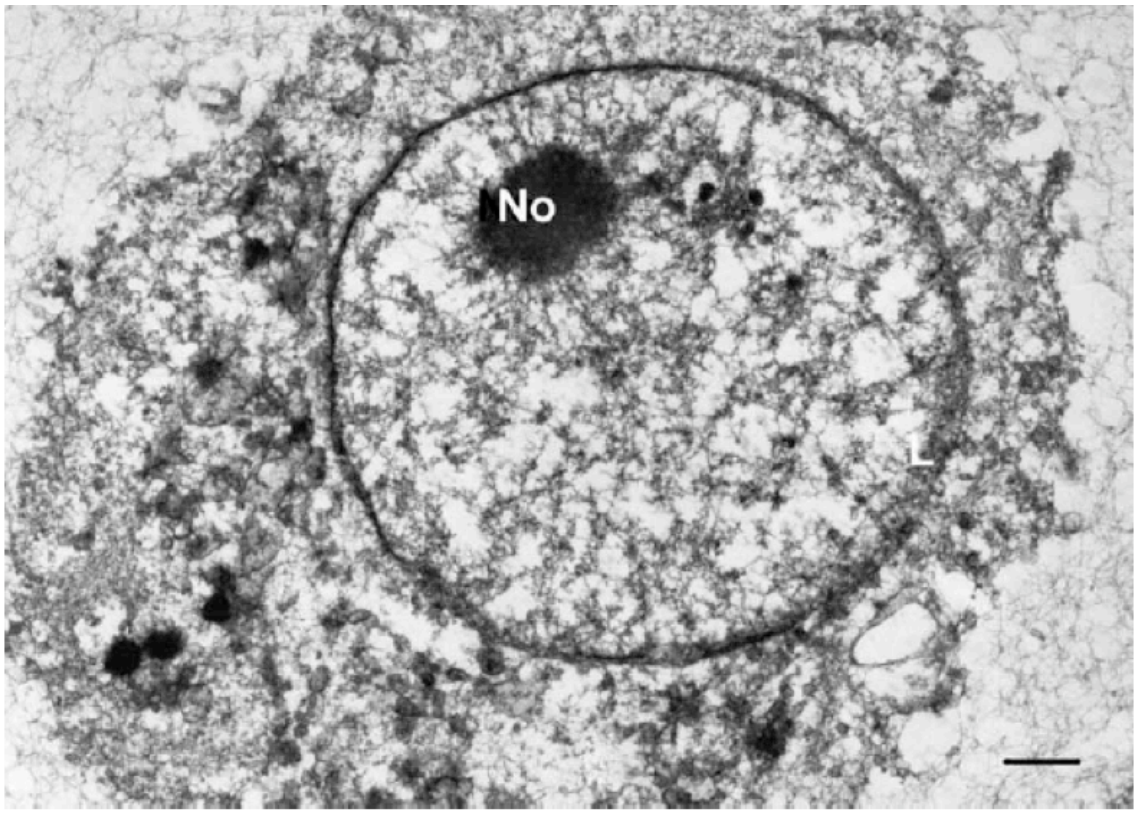

Figure 1.4 The nuclear matrix in a mammalian cell.

This electron micrograph of a HeLa cell shows the nucleus demarcated by a dense nuclear lamina (L) and a heavily stained nucleolus (No). Prior to this micrograph being taken, the cell was treated with nucleases to remove approximately 90% of its DNA. The nuclear matrix is seen here as an intricate network of filaments extending throughout the nucleus (Jackson et al., 2003).

1.2.2 Proteins of the nuclear matrix

The composition of the nuclear matrix can vary depending on the cell type and organism. In matrices prepared from different human cell lines and tissues, approximately 200-300 proteins are seen on two-dimensional gels (Fey and Penman, 1988; Gerner et al., 1998). Many matrix proteins are common to different cell types, including lamins, scaffold attachment factor A (SAFA), scaffold attachment factor B1 (SAFB1), and topoisomerase II (Fey and Penman, 1988; Mattern et al., 1997; Stuurman et al., 1990).

Lamins provide shape and stability to the nucleus and contribute to a wide range of nuclear functions. They are composed of a long α-helical rod flanked by globular N-terminal and C-terminal domains, and have the capacity to self-assemble into higher-order structures (Herrmann and Foisner, 2003; Stuurman et al., 1998). Lamins form type V intermediate filaments, which are the main constituents of the nuclear lamina, but they are also found in a diffused form throughout the nuclear interior (Dechat et al., 2008; Goldman et al., 2002). Most insights into the function of lamins have been obtained by studying cells and organisms with mutated lamin genes. While there is firm evidence that lamins are important in the regulation of transcription, DNA replication and repair, their precise roles in these processes remain to be elucidated (Kennedy et al., 2000; Liu et al., 2005; Moir et al., 1994; Spann et al., 2002). Cells with mutated lamins usually have irregularly shaped nuclei and abnormal positions of CTs. Furthermore, heterochromatin appears to be displaced from the nuclear periphery. Therefore, lamins are likely to maintain the integrity of nuclear architecture (Meaburn et al., 2007; Muchir et al., 2003; Nikolova et al., 2004; Sullivan et al., 1999). In humans, mutations in lamin genes cause a group of disorders called laminopathies, which include severe diseases such as Emery-Dreifuss muscular dystrophy, dilated cardiomyopathy, premature aging and progeroid syndromes (Bonne et al., 1999; Dechat et al., 2008; Worman and Bonne, 2007).

The nuclear matrix proteins SAFA and SAFB1 are associated with transcription. Both proteins contain the highly conserved SAF-box, a homodomain-like motif, which can interact with DNA sequences (Kipp et al., 2000; Oesterreich, 2003). Chromatin immunoprecitation experiments indicate that SAFA binds to DNA and recruits the transcriptional co-activator, p300 (Martens et al., 2002). This interaction is

accompanied by local histone acetylation, suggesting that SAFA might poise genes for transcription. By binding to DNA as well as the C-terminal of RNA Pol II and RNA processing factors, SAFB1 is thought to co-ordinate transcription by bringing these complexes into physical proximity (Nayler et al., 1998; Oesterreich, 2003; Weighardt et al., 1999). SAFB1 knockout mice show high pre-neonatal and neonatal lethality while their embryonic fibroblasts fail to undergo senescence and exhibit spontaneous immortalisation, highlighting the importance of this protein in development and cellular growth (Dobrzycka et al., 2006; Ivanova et al., 2005).

Topoisomerases alter the topological status of DNA to ensure correct progression of transcription, DNA replication, recombination and condensation. Topoisomerase II is one of the abundant components of the nuclear matrix (Berrios et al., 1985). In metaphase cells, it is mainly concentrated along the longitudinal axis spanning the chromatid, while at interphase it is diffused throughout the nuclear interior (Earnshaw et al., 1985). By catalysing the decatenation of DNA by an ATP-dependent mechanism, which cleaves and rejoins single- or double-stranded DNA, topoisomerase II relieves the torsional stress due to DNA supercoiling (Martins and Krawetz, 2007; Wang, 2002). Althought the role of topoisomerase II in nuclear structure is still uncertain, its interaction with chromatin at the bases of chromatin loops may play an important role in maintaining the integrity of chromosome architecture (Adachi et al., 1989; Earnshaw et al., 1985).

Other molecules identified in the nuclear matrix include components of the transcriptional machinery such as the phosphorylated form of RNA Pol II, HATs, HDACs, chromatin remodelling factors and numerous general transcription factors (Boulikas, 1995; Davie and Hendzel, 1994; Nayler et al., 1998; Patturajan et al., 1998; Reyes et al., 1997). The nuclear matrix also harbours molecules involved in mRNA processing including splicing factors and the ubiquitous heterogeneous nuclear ribonucleoproteins (hnRNPs), which are important in RNA biogenesis (Blencowe et al., 1994; Gerner et al., 1998; Mattern et al., 1996; Nayler et al., 1998). Furthermore, active DNA polymerase, primer recognition proteins (PRPs) and a number of other critical proteins involved in DNA replication are retained in the nuclear matrix (Collins and Chu, 1987; Jackson and Cook, 1986; Vishwanatha et al., 1992). The nuclear matrix is thus associated with fundamental nuclear processes.

1.2.3 The function of the nuclear matrix

The nuclear matrix has been hypothesised to determine the spatial arrangement of transcription and DNA replication by providing the structural framework for these processes (Berezney, 2002). Transcription and replication are highly compartmentalised in the nucleus **(**Fig 1.5**)**. By incorporating bromo-UTPs to newly transcribed mRNA, Jackson et al. (1993) showed that RNA synthesis occurs at a few hundred foci termed transcription factories (Jackson et al., 1993). Subsequent studies showed that each factory was a transcription 'hot spot' containing transcription factors and several active forms of RNA Pol which are involved in transcription initiation and elongation (Cook, 1999; Grande et al., 1997; Iborra et al., 1996; Pombo et al., 1999). Distinct transcriptional microenvironments are created in transcription factories that contain different sets of transcription factors and transcribe certain sets of genes (Xu and Cook, 2008). This organisation is likely to ensure effective interactions between components of the transcription machinery (Misteli, 2007). Furthermore, sharing a transcription factory could facilitate the co-ordinate transcription of commonly regulated genes.

There are several lines of evidence suggesting that transcription factories are attached to the nuclear matrix. Treatments of nuclei with hypertonic solutions show that as well as the main components of the transcriptional machinery, the nuclear matrix retains nascent RNA transcripts (Cook, 1999). As transcription factories are formed prior to transcription, they are likely to exist as independent sub-nuclear compartments (Osborne et al., 2007). Furthermore, the movement of molecules visualised during transcription reveals that upon activation, genes migrate to pre-existing relatively immobile transcription factories (Mitchell and Fraser, 2008).

As found for transcription, DNA replication is clustered at discrete sites known as replication factories, which can be visualised by incorporation of dUTP into newly replicating DNA (Cook, 2002). These foci contain the entire replication machinery, multiple replication origins as well as factors involved in chromatin assembly (Cook, 1999).

The binding of replication factories to the nuclear matrix is more controversial than for transcription factories. The nuclear matrix contains nascent DNA fragments, DNA Pol and other factors involved in replication (Cook, 1999). Furthermore, replication sites

coincide with intranuclear lamin structures, and initiation of replication is impaired when lamins are disrupted (Kennedy et al., 2000). However, replication factories are not formed prior to replication and appear to be maintained by the process of active DNA replication, suggesting that they are not independent sub-nuclear compartments (Kitamura et al., 2006).

A different view is that the global organisation of nuclear architecture, and therefore also of these essential nuclear processes, is not determined by the nuclear matrix but rather by the self-organising properties of the nuclear components (Misteli, 2005). In this model, the ongoing nuclear processes and the various affinities of chromatin for nuclear regulatory components are the major factors responsible for shaping the nuclear compartments (Cook, 2002; Misteli, 2001; Pombo and Branco, 2007). The global changes observed in nuclear architecture upon interference with transcription and replication support the self-organising nature of the nucleus (Lamond and Spector, 2003).

The major impediment in experimental testing of self-organising systems is the difficulty in separating structure from function. Computer simulations, which are based on a combination of data including the movement and the affinities of different nuclear components, are being developed to construct detailed models and make testable predictions (Misteli, 2007). The self-organising model does not exclude the existence of an underlying nuclear structure such as the nuclear matrix, as nuclear machineries can still be formed on a stable scaffold functioning as a landing platform (Gorski and Misteli, 2005; Misteli, 2007).

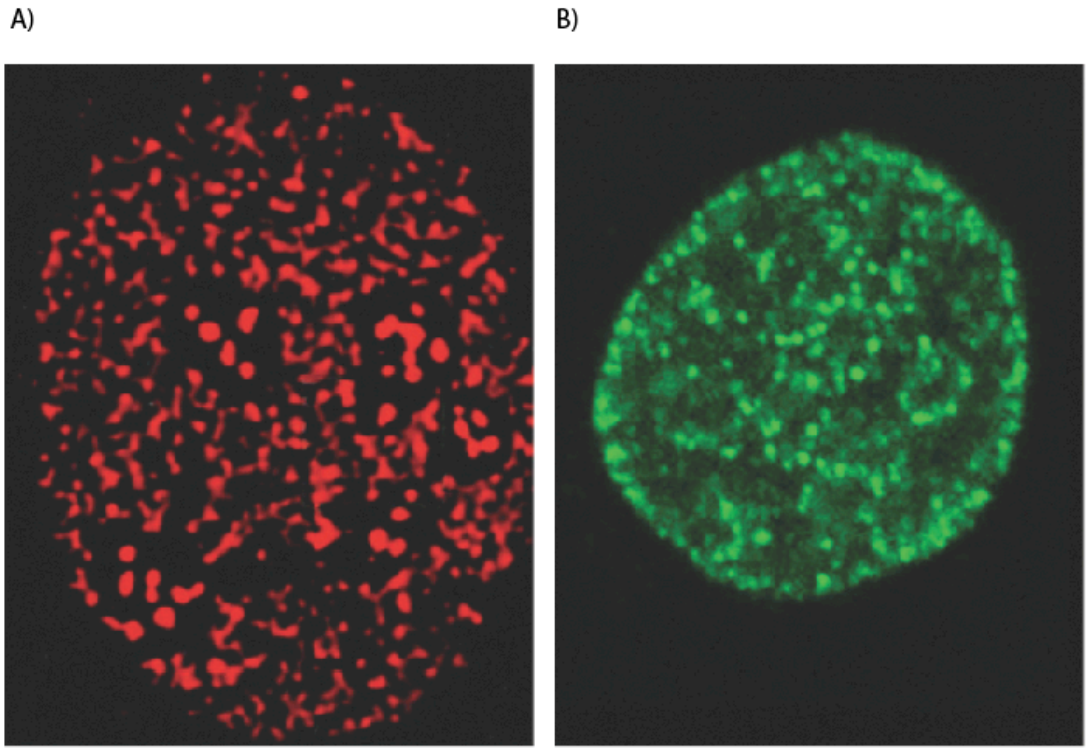

Figure 1.5 Distribution of transcription and replication factories throughout the nuclear interior.

A) Transcription sites were visualised by extending nascent transcripts in the presence of bromo-UTPs. Newly made RNA (red) is concentrated in discrete foci, called transcription factories. **B)** Replication sites were visualised by growing cells in the presence of bromo-dUTPs. Newly synthesised DNA (green) is clustered at sites known as replication factories (Misteli, 2007).

1.2.4 The controversy of the nuclear matrix

Despite evidence of an association between the nuclear matrix, nuclear architecture and different nuclear processes, the existence of a nuclear matrix as a separate nuclear entity remains controversial (Jackson, 2003; Pederson, 2000). Opponents of the nuclear matrix consider it to be an aggregation phenomenon generated by the preparative procedures, which can alter the structure of chromatin as well as the global nuclear environment. The removal of nucleic acids from the nucleus could disrupt DNA-protein interactions and allow the cationic sites of proteins previously bound to DNA to interact with other proteins (Pederson, 1998). Furthermore, different ionic conditions affect van der Waals' forces which in turn influence interactions between proteins, and could lead to the formation of artifactual protein complexes (Pederson, 1998).

On the other hand, the existence of a nuclear matrix is supported by the preservation of the functional domains of the nucleus such as splicing sites, PML bodies, CTs as well as replication and transcription factories within the nuclear matrix (Berezney, 2002). Since it is difficult to predict the changes caused by hypertonic buffers, methodologies have been developed to isolate the nuclear matrix in more physiological conditions (Jackson and Cook, 1986; Jackson et al., 1990; Mirkovitch et al., 1984). For example, cells can be encapsulated in agarose spheres and, after treating them with detergents and restriction enzymes, the chromatin and the soluble products is electro-eluted rather than removed with high-salt buffers (Jackson and Cook, 1986). Subsequent observations revealed that the ultrastructural features of these matrix preparations were similar to those obtained from extraction with high-salt buffers (Capco et al., 1982; Jackson and Cook, 1988). However a comprehensive analysis of the proteins constituting nuclear matrices isolated using different extraction methods is still lacking.

Based on the findings that nuclear matrix proteins such as special AT-rich sequence binding protein 1 (SATB1) are able to form intranuclear networks *in vivo*, there is now a growing recognition of the nuclear matrix as a fraction containing functional components of the eukaryotic nucleus that are necessary for the organisation of chromatin domains and the maintenance of sub-nuclear compartments (Cai et al., 2003; Oegema et al., 1997).

1.2.5 Features of matrix attachment regions

MARs, also described as genomic anchors, are short DNA sequences of approximately 100-2000 bp that tether chromatin loops to the nuclear matrix (Bode et al., 2003; Frisch et al., 2002; Ottaviani et al., 2008b; Singh et al., 1997). The strong interaction of MARs with the insoluble proteins of the nuclear matrix protects these sequences from nuclease digestion and treatments with high-strength ionic buffers (Heng et al., 2004). Subsequent digestion with proteases and RNase A followed by DNA extraction protocols allows MARs to be isolated and characterised.

MARs are ubiquitous features of eukaryotes since they organise the genome into hierarchical series of loop domains (Farache et al., 1990). The human genome has been estimated to contain between 30,000 and 80,000 MARs (Linnemann et al., 2007). So far, only a few hundred MAR sequences have been described. The best-characterised are those at the chicken lysozyme and α-globin, and the human β-globin and protamine loci (Farache et al., 1990; Girod et al., 2005; Martins and Krawetz, 2007; Ostermeier et al., 2003; Phi-van et al., 1998).

Although a consensus sequence has not yet been identified, MARs are rich in AT and repetitive sequences, and map to regions where the DNA is intrinsically curved or kinked and has a propensity for base unpairing (Bode et al., 1992; Fiorini et al., 2006; Kohwi-Shigematsu and Kohwi, 1990; von Kries et al., 1990). They are associated with topoisomerase I and II binding and cleavage sites, replication origins, transcription factor binding sites, nucleosome-free stretches, and DNase I cleavage sites, one of the hallmarks of 'open' chromatin (Boulikas, 1993; Glazko et al., 2003; Singh et al., 1997).

MARs are bound to the nuclear matrix either constitutively or transiently (Jackson et al., 1996; Kieffer et al., 2002; Liebich et al., 2002). Constitutive MARs are likely to maintain the higher-order chromatin structure of interphase and metaphase chromosomes (Glazko et al., 2003; Rollini et al., 1999). The transient, more dynamic associations are instead dependent on the transcription/replication status of the genomic domain, indicating a regulatory role (Heng et al., 2004).

MARs are often referred to as a distinct class of *cis*-acting elements as there is strong evidence linking them with the spatial and temporal regulation of gene expression and

DNA replication as discussed below. Furthermore, by dynamically organising chromatin in three-dimensional space, these genomic anchors play a prominent role in cellular development and disease (Cai et al., 2006; Iarovaia et al., 2004; Petrov et al., 2006).

1.2.6 Matrix attachment regions and transcription

The tethering of DNA to the nuclear matrix plays a vital role in transcription (Eivazova et al., 2007; Geyer, 1997; Heng et al., 2004). Using T cell differentiation as an example, the facilitation of transcription by MARs and their shaping of chromatin architecture to insulate chromatin domains from the effects of flanking chromatin will be described.

Upon stimulation by antigen, naive CD4+ helper T cells differentiate into effector Th1 and Th2 cells. In mice, *Ifng* (the gene for the cytokine IFN-γ) is silenced in naive T cells but transcribed in activated Th1 cells. The architecture of the *Ifng* locus has been analysed in these two cell types by a combination of 3C (see Section 1.1.9) and microarray technology (Eivazova et al., 2007). In naive T cells, *Ifng* was found to exist in a linear conformation, but in Th1 cells it is present in a chromatin loop, due to tethering of DNA to the nuclear matrix by MARs 7 kb upstream and 14 kb downstream of the locus. The absence of this selective DNA attachment to the nuclear matrix in naive T cells suggests that dynamic DNA anchors mediate the formation of the looped structure and the expression of the *Ifng* locus (Eivazova et al., 2007).

The molecular mechanisms by which MARs reorganise higher-order chromatin structure have been investigated in detail at the mouse Th2 cytokine locus, which contains a cluster of co-ordinately regulated genes, *Il4*, *Il13* and *Il5,* in a region of about 120 kb (Loots et al., 2000). These genes are silent in naive T cells but expressed in Th2 cells. Following activation of Th2 cells, expression of the nuclear matrix protein SATB1 is rapidly induced, and MARs within the locus mediate the formation of small loops by anchoring the loops onto a common protein core associated with SATB1 (Cai et al., 2006). Down-regulation of SATB1 expression by RNA interference prevents both the formation of this looped structure and transcriptional activation of the locus (Cai et al., 2006). In SATB1-null thymocytes the expression of many genes is spatially and temporally misregulated, and T cell development in SATB1-deficient mice is

prematurely blocked. These findings indicate that the binding of SATB1 at MARs regulates the expression of T cell differentiation genes by reorganising higher-order chromatin architecture (Alvarez et al., 2000; Cai et al., 2003). A similar MAR-mediated loop-formation mechanism regulates expression of the human β-globin gene cluster (Ostermeier et al., 2003; Wen et al., 2005).

Cai et al. (2003) reported that, at MARs, SATB1 recruits chromatin remodelling enzymes, which are able to either activate or repress the expression of adjacent sequences (Cai et al., 2003). Other studies have shown that MARs interact dynamically with basal components of the transcription machinery and with splicing factors (Donev et al., 2002; Rajaiya et al., 2006). As MARs associate with components of transcription factories as well as the nuclear matrix, it is tempting to speculate that their dynamic interactions with the matrix bring together proximal and distal regulatory DNA sequences and localise them close to transcription factories, thus promoting efficient regulation of gene expression (Fig 1.6).

Many genes are shielded by so-called 'insulator' elements from stimulatory or repressive transcriptional effects attributable to the chromatin state and regulatory elements in flanking regions. MARs commonly map to sequences flanking genes, and co-localise with some of the most extensively analysed insulator elements, including the *gypsy* retrotransposon in *Drosophila melanogaster*, suggesting that MARs have an insulator function (Nabirochkin et al., 1998). In *Drosophila melanogaster*, the nuclear matrix protein Su(Hw) binds to *gypsy*, creating chromatin loops (Byrd and Corces, 2003). Certain mutations in Su(Hw) that disrupt the loop structures render the insulator non-functional (Byrd and Corces, 2003; Valenzuela and Kamakaka, 2006). This suggests that the tethering of MARs to the nuclear matrix topologically constrains the DNA into looped structures, protecting the intervening DNA from the influence of *cis*-regulatory elements outside the loop. In vertebrates, CTCF, a ubiquitous nuclear matrix protein that binds to insulators, has also been shown to interact with MARs (Yusufzai and Felsenfeld, 2004). Experiments in a wide variety of higher eukaryotes have shown that in stably transfected cells, MAR-containing transgenes were expressed at higher levels compared with transgenes lacking MARs, indicating that the MARs shield the transgenes from the effects of the neighbouring chromatin regions (Girod et al., 2007; Halweg et al., 2005).

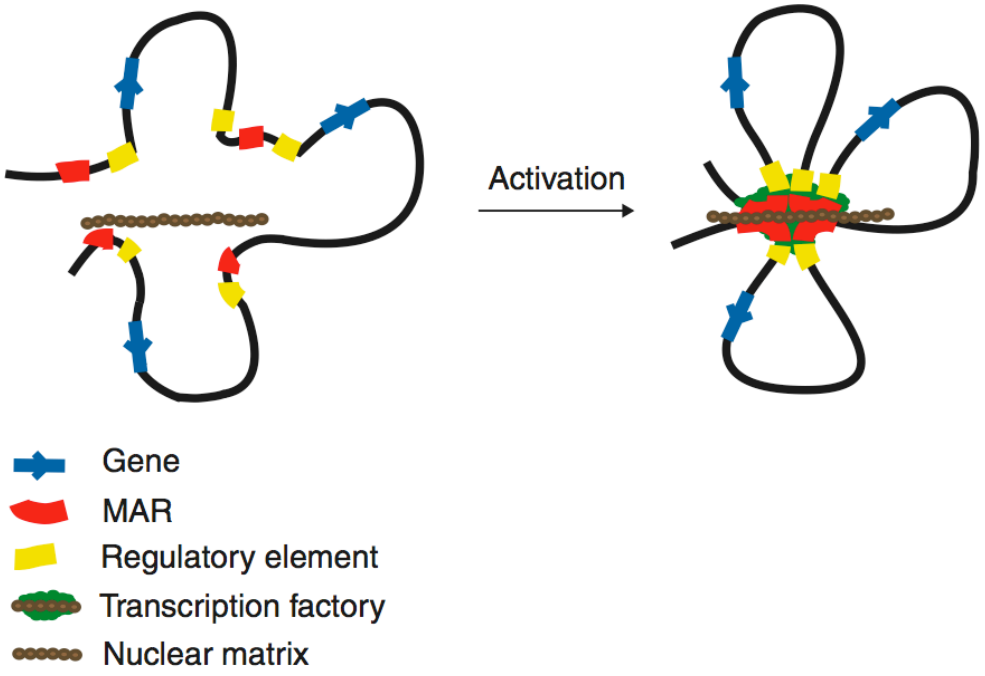

Figure 1.6 A simplified model depicting the function of matrix attachment regions (MARs) in gene regulation.

Activation of transcription is accompanied by the anchoring of MARs to the nuclear matrix. This results in the formation of an anchored chromatin loop that is insulated from the stimulatory or repressive effects of the flanking chromatin. The transcription machinery is assembled at the site of the MAR-nuclear matrix attachments. Interaction of MARs with the nuclear matrix brings together gene coding sequences, regulatory DNA elements and the transcription machinery, thus enabling specific genes to be co-ordinately regulated.

1.2.7 Matrix attachment regions and replication

To ensure that the genome is copied accurately, and only once per cell cycle, eukaryotes have evolved intricate mechanisms to regulate DNA replication. Some of the best-characterised origins of replication (ORIs) have been mapped to AT-rich genomic regions with base-unpairing elements. Futhermore, sequences at or near the ORIs for the human lamin B2 gene, the chinese hamster dihydrofolate reductase-β and -β' genes, the human β-globin gene, the chicken α-globin and lysozyme genes, and the *Xenopus* and mouse *c-myc* genes, function as dynamic MARs during the cell cycle (Djeliova et al., 2001a; Djeliova et al., 2001b; Girard-Reydet et al., 2004; Lagarkova et al., 1998; Mesner et al., 2003; Phi-van et al., 1998; Razin et al., 1991).

DNA replication is temporally and spatially ordered in the nucleus so that series of replicons are co-ordinately replicated at foci in the S-phase nucleus (Jackson and Pombo, 1998; Ma et al., 1998). Evidence that replication foci are associated with the nuclear matrix came first from EM (Hozak et al., 1993). Further support came from a study of nuclear matrix structures where DNA synthesis occurred at replication sites that were indistinguishable from those found in intact cells (Nakayasu and Berezney, 1989). In another study, DNA replication was found to be initiated at discrete chromosomal sites attached to the nuclear matrix (Radichev et al., 2005).

At replication foci, the nuclear matrix houses factors necessary for DNA replication, such as DNA Pol, the sliding clamp (PCNA) and single-strand binding protein (RPA), and provides structural support throughout the replication process. Wu and Gilbert (1996) proposed that origins are selected and replicon size determined in early G1-phase of the cell cycle (Wu and Gilbert, 1996). Using an *in vitro* system, it was subsequently shown that MCM2, a component of the pre-replicative complex, is loaded onto chromatin gradually and cumulatively throughout G1-phase, but is rapidly excluded from active replication foci in S-phase (Dimitrova et al., 1999). Tatsumi et al. (2003) reported a similar cycle of events for ORC1, a component of the replication initiation complex at ORIs (Tatsumi et al., 2003). This coincides with the recruitment of the chromatin-bound ORC2-5 complex to a structure likely to be the nuclear matrix, suggesting a link between the accumulation of ORC1 and the assembly of the replication complex in human nuclei (Ohta et al., 2003).

These observations fit a model in which MARs stably anchor the replicon ends and, during G1-phase, small-scale sub-chromosomal chromatin refolding recruits ORIs to the nuclear matrix, where factors accumulate to form the pre-replicative complexes (Fig 1.7). Subsequently, as ORIs begin to replicate in S-phase, certain protein factors dissociate from the chromatin and undergo proteolysis as part of a control mechanism to prevent re-replication, thus releasing the ORIs from the nuclear matrix. In the meantime, replication continues at the initial location as DNA is reeled through the replication machinery or replication factory (Hozak et al., 1993). At the ends of replicons, stable MARs could act as barriers to prevent the accumulation of supercoiled DNA structures, while providing binding sites for topoisomerase II to resolve replication intermediates.

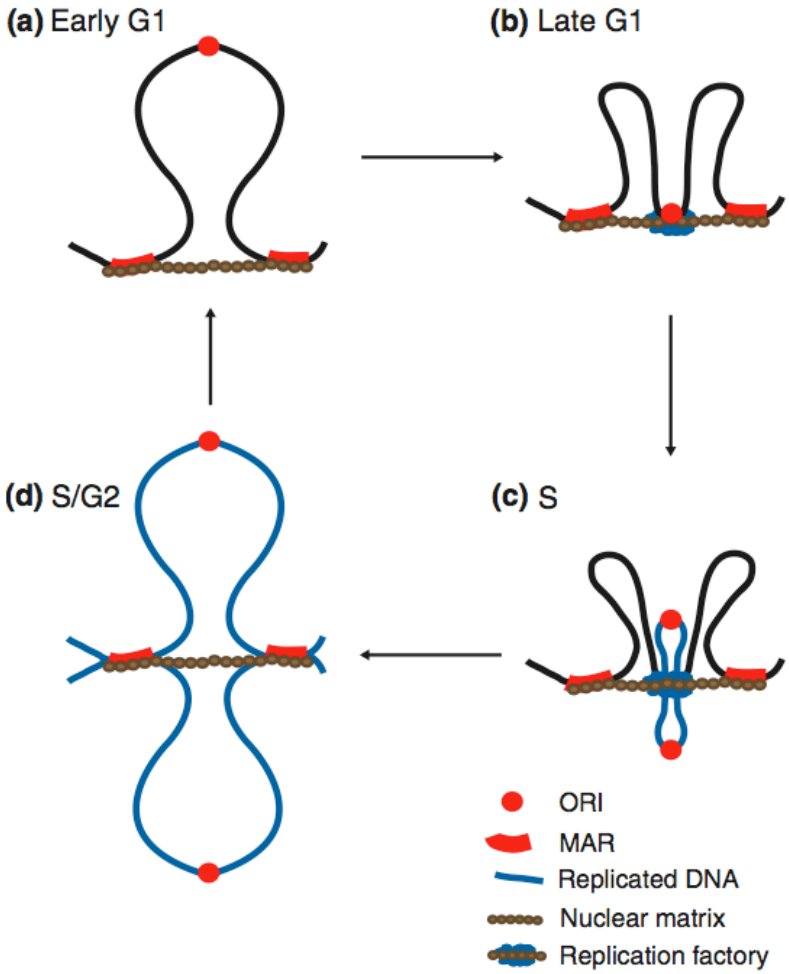

Figure 1.7 DNA replication is organised at the nuclear matrix.

(a) Replicons are defined in early G1-phase of the cell cycle by attachment of MARs to the nuclear matrix. **(b)** In late G1-phase, origins of replication (ORIs) are recruited to the nuclear matrix and replication factors assemble at these sites, licensing the chromatin for replication. **(c)** Once the appropriate mitogenic stimuli have been received, cells enter S-phase, at which ORIs become activated. Following initiation of replication at a particular locus, the two identical newly replicated ORIs probably dissociate from the nuclear matrix. Two loops of replicated DNA gradually emerge (shown in blue), while the yet-to-be replicated DNA of the replicon moves through the replication factory. **(d)** At the end of S-phase, the replication machinery is dismantled. Adapted from (Anachkova et al., 2005), (Ottaviani et al., 2008b).

1.2.8 Matrix attachment regions and disease

Integration of retroviral DNA into the host genome is essential for viral replication. Although retroviral integration sites lack a consensus sequence, they are often AT-rich and have base unpairing and DNA-bending and unwinding elements (Rampalli et al., 2003; Shera et al., 2001). DNA sequence analysis indicates that both DNA tumuor viruses and retroviruses integrate within or close to MARs (Fig 1.8) (Johnson and Levy, 2005; Kulkarni et al., 2004). Furthermore, the efficiency of transcription of the retrovirus HIV-1 is determined by the proximity of its integration to MARs (Rampalli et al., 2003). As SATB1 binds to MARs flanking HIV-1 integration sites and silencing of SATB1 gene expression alters the pattern of integration sites, it has been suggested that retroviruses use MARs to form viral pre-integration complexes (Kumar et al., 2007b).

MARs also appear to play a role in the development of some cancers. Chromosome rearrangements are hallmarks of certain malignancies and inherited genetic disorders. The breakpoints of recurrent translocations in leukaemia as well as deletions involving the breast-cancer susceptibility genes *BRCA1* and *BRCA2* occur at MARs, indicating that the bringing together of these sequences at the nuclear matrix facilitates their illegitimate recombination (Iarovaia et al., 2004; Welcsh and King, 2001). Patients who develop leukaemia following treatment of a primary tumour with inhibitors of topoisomerase II often have specific chromosome translocations in their cancer cells whose breakpoints contain MARs, emphasising the importance of the chromatin environment in the generation of chromosome aberrations (Strick et al., 2006).

Fragile sites are hypervariable regions that generate genomic instability in tumours. Certain fragile sites contain long AT-rich minisatellites, called AT-islands, which function as MARs (Jackson et al., 2003). AT-islands are susceptible to considerable repeat expansion, which, in the fragile site FRA16B associated with leukaemia, appears to strengthen their attachment to the nuclear matrix (Jackson et al., 2003). The presence of abnormal transcripts of the tumour suppressor gene *WWOX* (which spans FRA16B) in the absence of detectable mutations or deletions may be caused by aberrant chromatin architecture due to enhanced MAR anchoring by expanded AT-islands (Woynarowski, 2004).

The identification of AT-islands has led to the emergence of a new class of drugs that specifically alkylate them (Woynarowski, 2004). These drugs exhibit an extraordinary cytotoxicity, which is likely to be due to their disruption of replication and transcription, the two essential nuclear processes organised at MARs (Fig 1.9). One of these drugs, bizelesin, binds specifically to the minor groove of DNA at AT-rich regions and generates interstrand cross-links. It has high cytotoxic activity *in vitro* towards a broad spectrum of human cancer cell lines and, more importantly, high activity against various tumours engrafted in mice (Alley et al., 2004; Carter et al., 1996). While extensive development will be needed to make these compounds safe anti-cancer drugs for clinical use, their DNA sequence-specificity might offer a novel approach for targeting tumour cells containing expanded AT-repeat sequences.

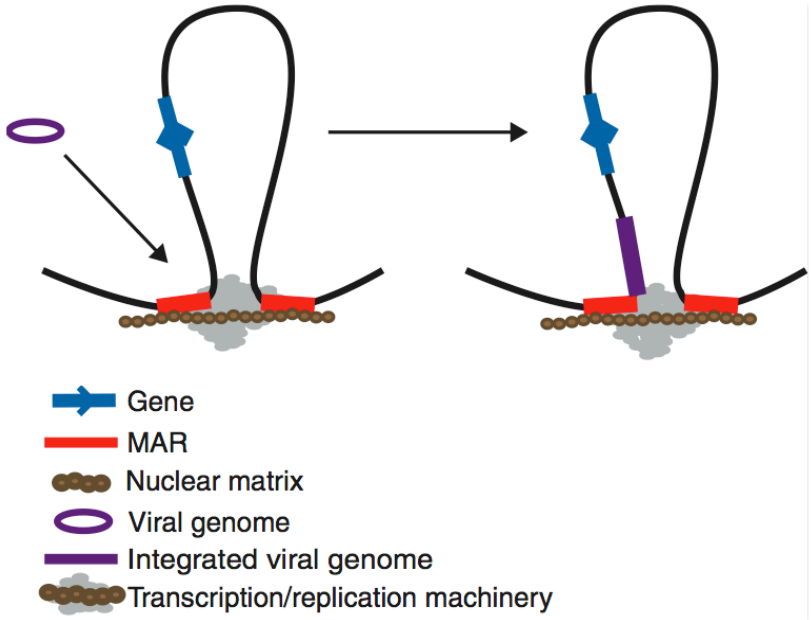

Figure 1.8 Schematic representation of viral genome integration.

Tumour viruses and HIV-1 integrate near MARs attached to the nuclear matrix, where the transcription and DNA replication machinery is assembled. The viral genome is thus integrated near the machinery required for its transcription and replication. Adapted from (Shera et al., 2001), (Ottaviani et al., 2008b).

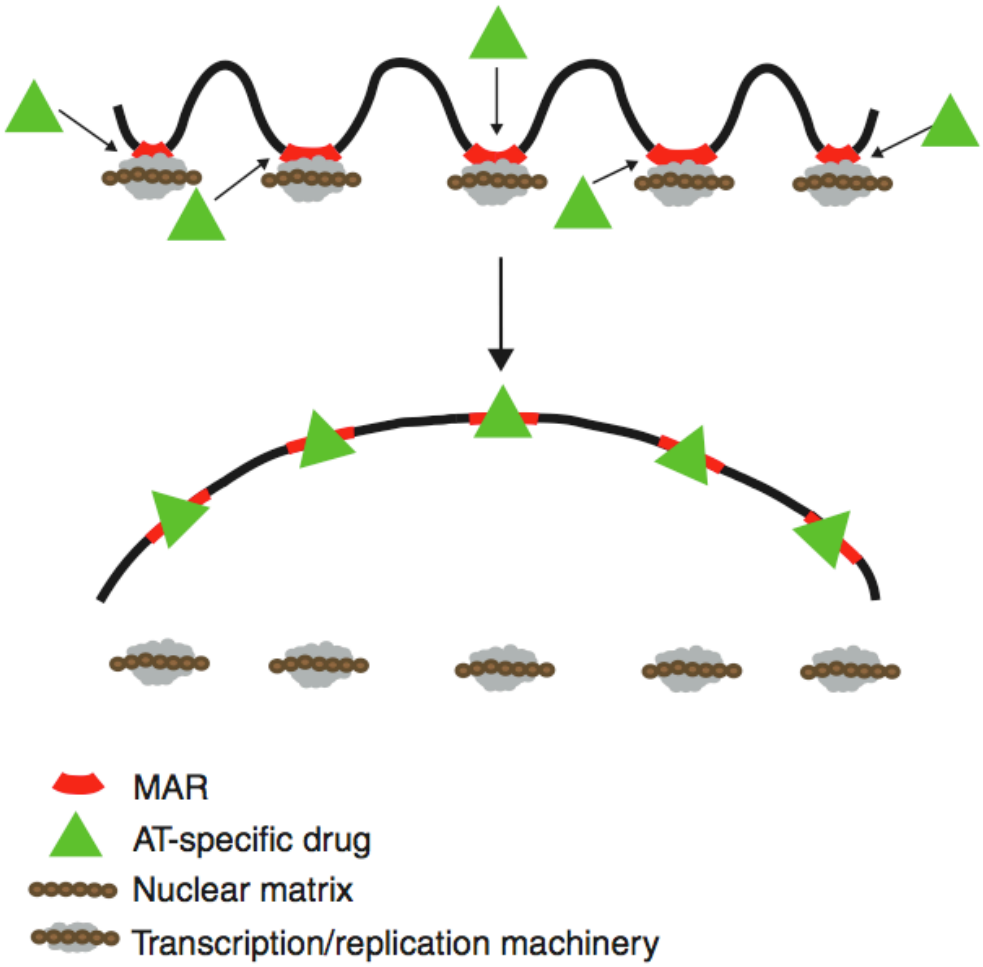

Figure 1.9 Proposed mechanism for the cytotoxic action of AT-specific drugs.

The drugs bind to AT-rich MARs in chromatin, cross-linking the two strands of the DNA. This leads to the disruption of processes such as transcription and DNA replication that are initiated at or in the vicinity of MARs (Ottaviani et al., 2008b).

1.3 The major histocompatibility complex

The human major histocompatibility complex (MHC) is one of the most important regions of the genome with respect to immunity to pathogens, autoimmunity and transplantation. It contains clusters of genes that are critical for innate and adaptive immune function, including antigen processing and presentation as well as the complement system. MHC genes are under tight regulatory control and mis-regulation of their expression gives rise to a range of diseases.

1.3.1 Overview

The human major histocompatibility complex (MHC) is one of the most extensively studied regions of the genome. In humans, it extends over approximately 4 Mb on the short arm of chromosome 6. The mouse MHC, called H2, was first identified and associated with immunity about 60 years ago by George Snell (Fernando et al., 2008). The locus was recognised in humans a few years afterwards where it was termed the MHC or human leukocyte antigen (HLA) region since MHC molecules were found to be present on the surface of leukocytes cells (Dausset, 1958; Payne et al., 1964).

The classical MHC is divided into 3 sub-regions, class II (centromeric), class III, and class I (telomeric) (Fig 1.10). Relative to other parts of the genome, the MHC is extremely gene-rich and highly polymorphic. The identification of MHC related genes beyond the boundaries of the classically defined MHC locus, has led to the description of the extended MHC, (xMHC). The xMHC spans about 7.6 Mb and includes the extended class I sub-region, located upstream of the MHC class I and the extended class II sub-region, located downstream of the MHC class II (Horton et al., 2004).

The genes contained in each class of the MHC have related functions and are subject to co-ordinate regulation. The products of the classical MHC class I and II genes are involved in antigen processing and presentation to T cells, while genes in the class III region encode complement components which are involved in the induction of inflammatory responses and damages to pathogens via proteolytic cleavages of glycoproteins.

Approximately 40% of the 224 loci identified within the MHC play a role in the immune system. HLA genes, which encode cell-surface antigen-presenting molecules, are the most polymorphic in the entire genome, some having over 200 alleles. *HLA* genes are closely linked and the entire MHC is inherited as a haplotype from each parent. Combinations of different *HLA* alleles give rise to an extremely large number of haplotypes, but in a population certain haplotypes occur more frequently than expected by chance (Choo, 2007). This phenomenon, which defines the non-random association between alleles at adjacent loci is known as linkage disequilibrium (LD). The high LD characterising the MHC could favour the persistence of favourable haplotypes in the population. Each person is heterozygous for *HLA* genes as the genome contains a paternal and maternal haplotype, which are co-dominantly expressed. This generates an extremely high level of genetic variation, which is required to mount an effective immune response against a wide range of evolving pathogens.

The MHC is associated with more diseases than any other region of the human genome (Price et al., 1999). These include autoimmune diseases such as multiple sclerosis, type 1 diabetes, systemic lupus, ulcerative colitis and rheumatoid arthritis (Fernando et al., 2008; Rubio et al., 2004; Thorsby, 1997). Although the exact mechanisms underlying the role of the MHC in these diseases are still unclear, association studies indicate that specific *HLA* haplotypes are related to autoimmune disease susceptibility (Fernando et al., 2008). As the MHC is critical in combating pathogens, MHC variants can confer susceptibility to infectious diseases such as malaria and HIV (Carrington et al., 1999; Hill et al., 1991). Furthermore, HLA molecules are important in transplantation, since they are recognised on the organ graft by host T cells causing graft rejection.

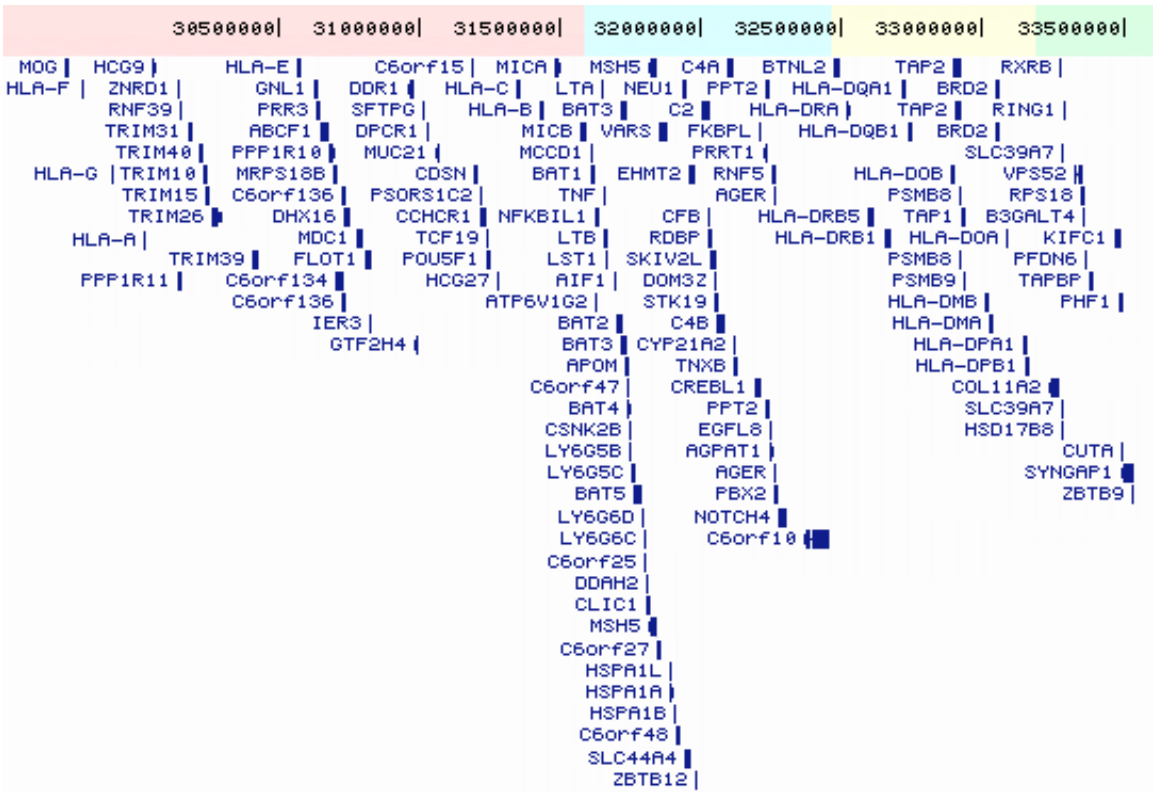

Figure 1.10 Gene map of the human MHC.

The human MHC spans approximately 4 Mb on the short arm of chromosome 6. Gene annotation is according to the RefSeq database (http://genome.ucsc.edu, May 2004 Assembly). The different classes of the MHC are highlighted in different colours: classical class I – red; classical class III – blue; classical class II – yellow; and extended class II – green.

1.3.2 MHC class I genes

The MHC class I region contains the classical *HLA-A*, *HLA-B* and *HLA-C* genes, which encode the α-chain (also called heavy chain) of the class I molecule. The β_2-microglobulin protein, encoded by a gene on chromosome 15, is non-covalently bound to the α-chain to form the complete heterodimeric structure of the class I molecule (Fig 1.11) (Bjorkman and Parham, 1990). The α-chain contains three extracellular domains (α-1, α-2 and α-3), a hydrophobic transmembrane domain and a cytoplasmic tail (Bjorkman et al., 1987). The α-1 and α-2 domains, constituted by variable amino acid sequences, form a unique groove accommodating the processed peptide antigen, which is typically 8-10 residues long (Choo, 2007).

The classical class I genes are expressed on all nucleated cells of the body and bind to fragments of peptides derived from intracellular parasites such as viruses or proteins normally synthesised by the cell (self-molecules). These peptides are degraded by the proteasome complex in the cytosol and then are actively translocated into the lumen of the endoplasmic reticulum (ER) by the membrane transporter TAP, a heterodimer constituted by the TAP1 and TAP2 subunits (Gromme et al., 1999; Kloetzel and Ossendorp, 2004; Powis et al., 1992; Powis et al., 1993; Shepherd et al., 1993). Once in the ER, these peptides are loaded onto the fully folded classical class I molecules and then transported via the secretory pathway to the cell surface where they interact with CD8+ cytotoxic T lymphocytes (CTLs) (Kloetzel and Ossendorp, 2004).

The MHC also contains the non-classical class I molecules, encoded by the *HLA-E*, *HLA-F* and *HLA-G* genes, which are evolutionary and structurally related to the classical class I molecules. The non-classical class I molecules also play a role in antigen presentation but are less polymorphic and exhibit a more restricted expression pattern than the classical class I molecules (Gobin and van den Elsen, 2000). While expression of *HLA-E* is ubiquitous, expression of *HLA-F* and *HLA-G* is mainly confined to B lymphocytes and extravillous cytotrophoblast cells, respectively (Gobin and van den Elsen, 2000). In addition to presenting peptides derived from the signal sequence of classical class I molecules to natural killer (NK) cells, it is emerging that HLA-E might also mediate CTL responses (Hoare et al., 2006). HLA-F has been undetected at the cell surface, however, structural predictions in combination with biochemical studies suggest it might be able to reach the cell surface by binding cell type specific peptides

and interact with immunoglobulin-like receptors ILT2 and ILT4 (Lepin et al., 2000). Although the function of HLA-G is unknown, there is evidence suggesting it can inhibit NK and T cell mediated cell-lysis and contribute to the normal progression of pregnancy and growth of the developing foetus (Hviid, 2006).

Additional genes with immune-related functions contained in the MHC class I region are *MICA* and *MICB* which encode membrane bound polypeptides different from the HLA class I proteins (Stephens, 2001). The MIC molecules do not associate with the β_2-microglobulin and the reduced sized of their groove cannot accommodate peptides for presentation to T cells (Stastny, 2006). Several studies indicate that MICA and MICB function as ligands for the receptor NKG2D, which is found on NK cells as well as different types of T cells. This interaction seems to calibrate the cytotoxic activity of these cells and may play a role in eliminating infected and tumour cells (Schrambach et al., 2007).

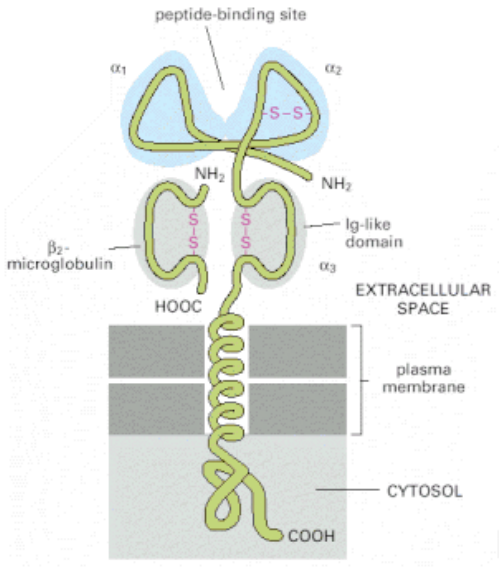

Figure 1.11 Diagram of the HLA-A, HLA-B or HLA-C molecule.

The α-chain has three extracellular domains, α_1, α_2 and α_3. It is non-covalently associated with the smaller polypeptide, β_2-microglobulin. The α_3-domain and β_2-microglobulin are Ig-like. While β_2-microglobulin is invariant, the α-chain is extremely polymorphic, mainly in the α_1-and α_2-domains, which form the peptide-binding site (Alberts et al., 1994).

1.3.3 MHC class II genes

The MHC class II region contains the classical *HLA-DP*, *HLA-DR* and *HLA-DQ* loci, which encode transmembrane proteins, each constituted by two non-covalently bound polypeptide chains: the α-chain and the β-chain (Fig 1.12). Each polypeptide chain has two extracellular domains (α-1, α-2; β-1, β-2), a single transmembrane and an intracytoplasmic domain (Jones et al., 2006). The N-terminal α-1 and β-1 domains form a groove, the peptide-binding domain, which binds to fragments of processed polypeptides. The α-2 and β-2 domains form an immunoglobulin-like domain (Stern et al., 1994).

The classical class II molecules are expressed by antigen presenting cells (APCs) such as dentritic cells, macrophages and B cells, but their expression can be induced in many cell types by cytokines including interferon-γ (IFN-γ). The antigens presented by the classical class II molecules are derived from peptides that are translocated from the extracellular space into the cytosol via micropinocytosis, phagocytosis or receptor-mediated uptake (Kumanovics et al., 2003).

Synthesis of the classical class II molecules takes place in the ER where they bind to the invariant chain (Ii) molecule at the peptide-binding domain. Ii molecules act as chaperones directing the classical class II molecules to the endocytic pathway and prevent the peptide grooves from binding with self-molecules (Cresswell, 1996). In the endocytic pathway, increasingly acidic endosome compartments degrade the internalised peptides to about 13-18 amino acids and the Ii molecule is gradually degraded to a small fragment known as class II associated invariant chain peptide (CLIP), which serve to stabilise the folded classical class II molecule (Bakke and Dobberstein, 1990). The non-classical class II molecules HLA-DM and HLA-DO mediate the removal of CLIP and the loading of the antigens onto the groove of the classical class II molecule, which is then transferred to the cell surface where it interacts with CD4+ T cells to regulate the adaptive immune responses against invading pathogens (Kumanovics et al., 2003).

Other genes in the MHC class II region encode molecules which participate in the class I antigen presentation pathway (Kumanovics et al., 2003). These include the transporter genes *TAP1* and *TAP2* (see Section 1.3.2) as well as the proteasome genes *PSMB9* and

PSMB8, which collectively form a gene cluster. Within the MHC, there are genes that are not related to antigen processing and presentation but seem to be involved in other aspects of the immune response. An example is *DAXX*, which encodes a molecule involved in growth arrest and/or apoptosis of auto-reactive lymphocytes (Kiriakidou et al., 1997; Muromoto et al., 2004).

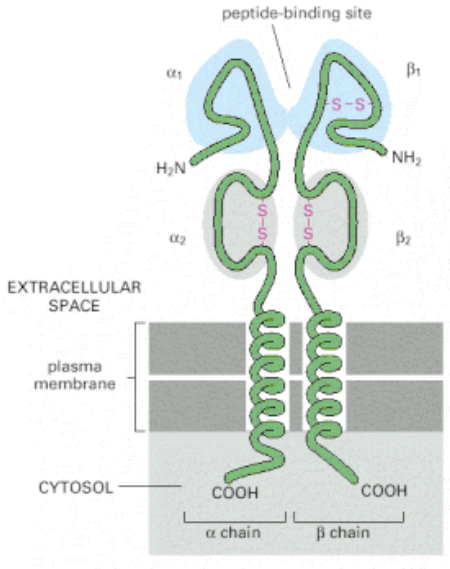

Figure 1.12 Diagram of the HLA-DP, HLA-DR or HLA-DQ molecule.

The α-chains and β-chains are polymorphic, mainly in the α_1- and β_1-domains, which interact to form a groove that binds peptide fragments. The α_2- and β_2-domains are Ig-like (Alberts et al., 1994).

1.3.4 MHC class III genes

Compared to the class I and II regions, which are characterised by numerous genes involved in antigen processing and presentation pathways, the class III region harbours a more heterogeneous collection of genes. These genes encode proteins that are crucial for various aspects of immunity including activation of the complement system, hormone synthesis, inflammation and cell stress, synthesis of lymphotoxins and heat shock proteins (Hauptmann and Bahram, 2004).

The complement system is a major effector of innate immunity and also interacts with the cells of the adaptive immune system to influence adaptive immune responses (Morgan and Harris, 2003). The molecules constituting the complement system include approximately 25 soluble proteins existing in the serum as inactive pro-enzymes (zymogens), which can be activated to augment antibody responses, lyse foreign cells and promote the clearing of immune complexes and apoptotic cells (Sim and Tsiftsoglou, 2004). The complement system can be activated by three different pathways known as the classical, lectin and alternative pathways.

The MHC class III region contains genes encoding two critical pro-enzymes, C4 and C2. The trigger of the 'classical pathway' is the binding of antibodies with their antigen (Sim and Tsiftsoglou, 2004). This interaction activates a component of the complement system C1, which cleaves C4 into the sub-components C4a and C4b, and C2 into the sub-components C2a and C2b (Sim and Tsiftsoglou, 2004). The binding of C4b to C2a forms the C3 convertase complex, a protease that initiates a cascade culminating in the formation of the membrane attack complex (MAC), which creates a transmembrane pore on the target cell leading to its lysis (Law and Dodds, 1997).

The C4 and C2 molecules are also involved in the 'lectin' pathway, which is activated by the binding of mannose-binding lectin to mannose residues on the surface of pathogens. This interaction activates the serine proteases MASP-1 and MASP-2, which in turn induces the cleavage of C4 and C2 into their active sub-components leading to pathogen elimination via the same steps described for the classical pathway (Holmskov et al., 2003).

The MHC class III region encodes an additional important pro-enzyme, the B-factor that activates the complement system through the 'alternative' pathway. This cascade is initiated by the spontaneous hydrolysis of C3, which then forms a complex with the B-factor. This process leads to the cleavage of the B-factor by the D protease and the formation of the C3bBb complex, which is analogous to the C3 convertase of the classical and lectin pathways (Sim and Tsiftsoglou, 2004). Once formed, the alternative pathway C3 convertase interacts with other proteins, including factor I and H, to regulate opsonisation.

The MHC class III region contains immune-related loci that are not part of the complement system. These include the tumour necrosis factor (*TNF*), lymphotoxin-α (*LTA*) and lymphotoxin-β (*LTB*) genes, which encode cytokine precursors involved in the signalling networks mediating innate and adaptive immune responses (Deakin et al., 2006). Disruption of these signalling pathways reduces inflammation as well as resistance to selected pathogens (Ware, 2005). Additional class III genes with important functions in immunity are *AIF1*, *LST1* and *NCR3* that play a role in inflammation, lymphocyte development and NK activation, respectively (Deininger et al., 2002; Pende et al., 1999; Rollinger-Holzinger et al., 2000).

1.3.5 Transcriptional co-regulation of MHC genes

The expression of classical MHC genes, as well as those involved in antigen processing and presentation, is subject to co-ordinate regulation. This provides cells with an effective system to orchestrate the interaction of functionally linked MHC molecules at the appropriate time and place. The co-regulation of MHC genes is complex and involves the activity of conserved genomic regulatory elements as well as epigenetic mechanisms.

The classical and non-classical MHC class I genes, with the exception of *HLA-G*, share three major regulatory elements: enhancer A, IFN-stimulated response element (ISRE) and the SXY module (Fig 1.13)(van den Elsen et al., 1998; van den Elsen et al., 2004). These sequences are localised within about 200 nucleotides upstream of the transcription start site and are part of the 5' promoter region (van den Elsen et al., 2004). Enhancer A contains binding sites for the nuclear transcription factor kB (NF-

kB), while ISRE interacts with the IFN regulatory factors 1 (IRF-1) and 2 (IRF-2) (Gobin et al., 1998; Gobin et al., 1999). The synthesis of these molecules, which modulate the induction and suppression of these class I genes, is regulated by cytokines including TNF and IFNs (van den Elsen et al., 2004). The SXY module, which contain the S, X (comprising X1 and X2 sequences) and Y motifs, binds to a multiprotein complex constituted by regulatory factor X (RFX), activating transcription factor (ATF), cAMP response element binding protein (CREB) and the nuclear factor Y (NFY) (van den Elsen et al., 2004). This complex forms a landing platform for the crucial transcription factor, class II Transactivator (CIITA), to up-regulate the transcription of classical and non-classical MHC class I genes (Krawczyk et al., 2008).

The SXY module is also part of the promoter region of the classical and non-classical MHC class II genes (Fig 1.14). In contrast to the class I genes, the class II genes do not have the enhancer A and ISRE sequences in their promoter regions. The absence of these *cis*-regulatory sequences could contribute to the differences in expression between class I and II genes observed in different cell types. Studies on cell lines with mutated forms of CIITA and RFX show that they are essential for the expression of class II genes and also contribute, albeit to a lesser extent, to transcription of class I genes (Krawczyk et al., 2008). Mutations in the *CIITA* or *RFX* genes are responsible for a severe hereditary immunodeficiency disease, the bare lymphocyte syndrome (BLS), which is characterised by lack of expression of the class II genes and down-regulation of class I genes (Reith and Mach, 2001). The residual expression of class I genes is due to their additional upstream regulatory elements which can partially compensate for the absence of functional RFX or CIITA. The expression of classical class II genes is also modulated by conserved S'-Y' modules, which function as transcriptional enhancers (Krawczyk et al., 2004). These motifs contain the binding sites for CIITA and RFX and are localised several hundred base pairs upstream of the *CIITA* transcription initiation sites (Krawczyk et al., 2004).

CIITA activates and regulates gene expression, not only by interacting with several transcription factors at the SXY and S'-Y' modules, but also by binding to chromatin remodelling enzymes (Zika and Ting, 2005). Several studies have shown that CIITA is responsible for recruiting HATs, HMTs, and the ATP-dependent remodelling factor Brahma-related gene 1 (BRG1) at selected promoter regions (Mudhasani and Fontes, 2002; Wright and Ting, 2006; Zika and Ting, 2005).

Expression of MHC class II genes in APCs is due to their constitutive expression of CIITA. The expression of class II can be activated in many other cell types by inducing the synthesis of CIITA with IFN-γ via the JAK-STAT signalling pathway (Darnell et al., 1994). The pathway starts with the binding of IFN-γ to specific receptors on the cell surface, which are associated with protein kinases belonging to the Janus kinase (JAK) family (Darnell et al., 1994; Hibino et al., 1992). Ligand attachment leads to the phosphorylation of JAKs, which can than activate latent cytoplasmic signal transducers and activators of transcription (STAT) proteins through phosphorylation (Lee and Benveniste, 1996). Subsequently, the activated STATs dimerise and rapidly translocate into the nucleus where they binding the IFN-γ activation sequence (GAS) elements present in the promoters of IFN-γ inducible genes, including *CIITA* (Fig 1.15) (Darnell et al., 1994).

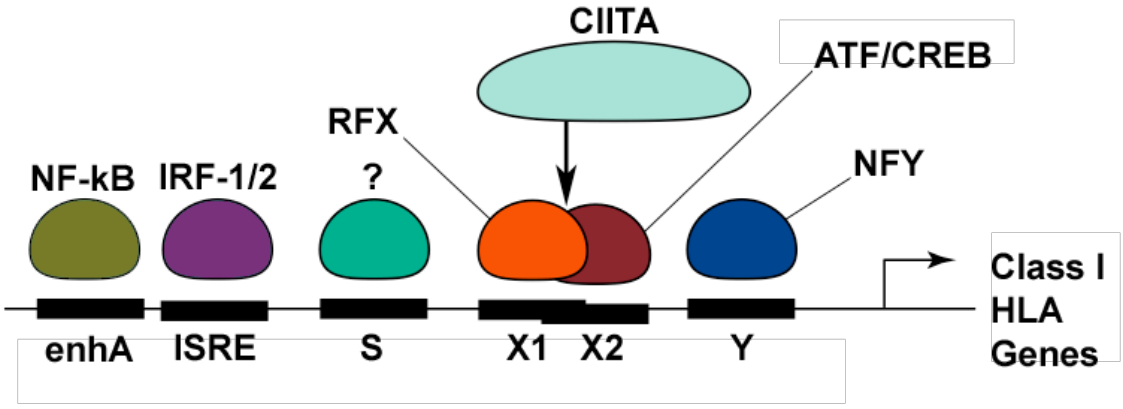

Figure 1.13 The conserved regulatory elements of the classical and non-classical class I *HLA* genes.

The SXY module is bound by RFX, ATF/CREB and NFY. CIITA exerts its activity primarily through the SXY module and its binding proteins. The enhancer (enhA) and ISRE are bound by NF-kB and IRF-1/2 factors, respectively. The protein binding the S module has not been identified.

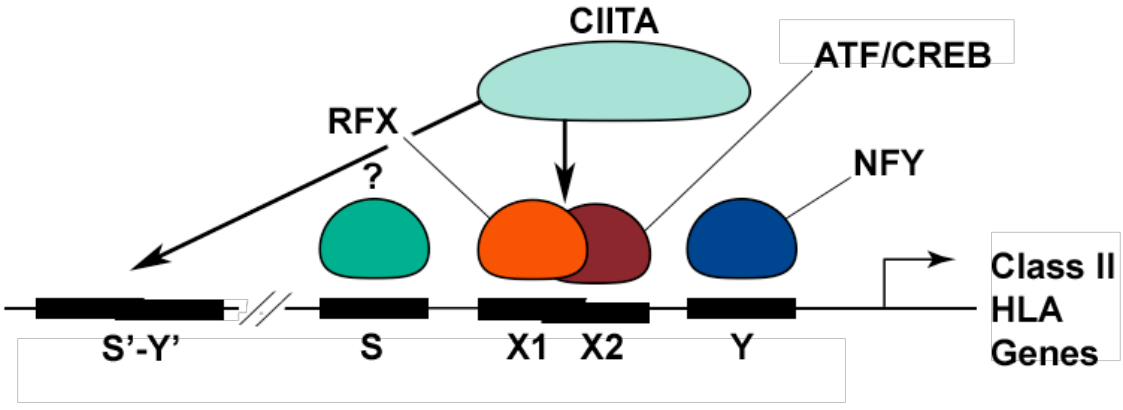

Figure 1.14 The conserved regulatory elements of the classical and non-classical class II *HLA* genes.

The SXY module is bound by RFX, ATF/CREB and NFY. CIITA exerts its activity through the SXY and S'-Y' modules.

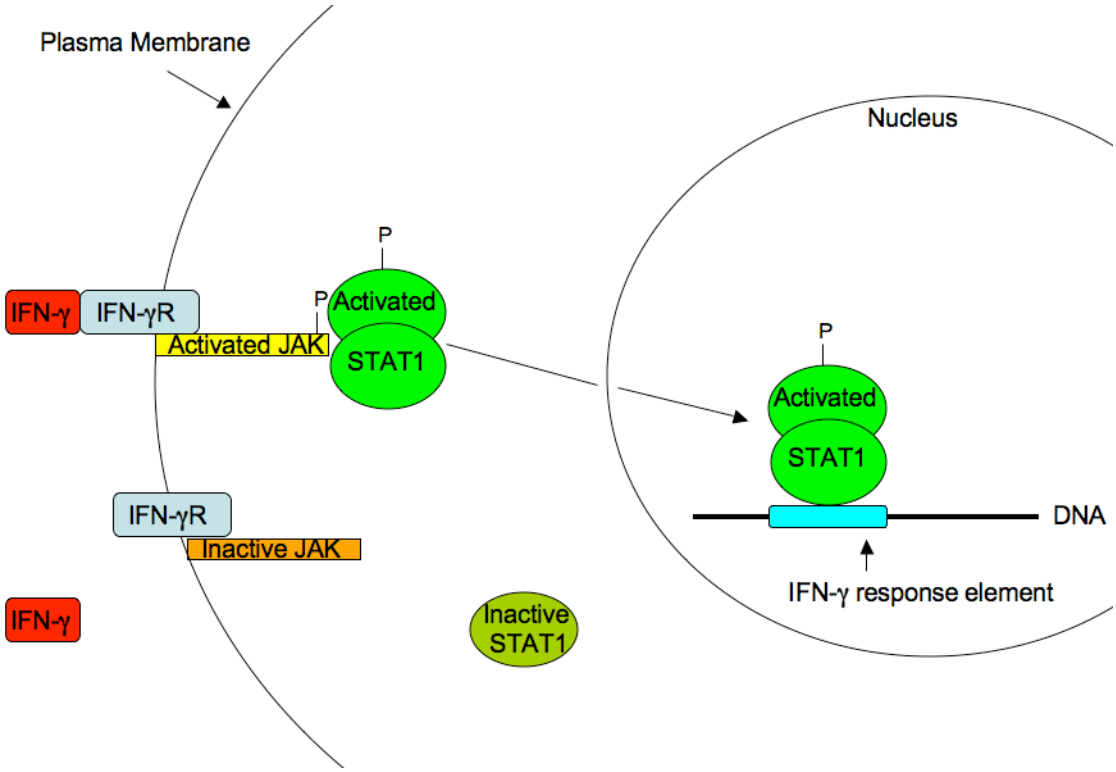

Figure 1.15 The JAK/STAT signalling pathway.

(1) Binding of IFN-γ to specific receptors on the cell surface activate protein kinases belonging to the JAK family. **(2)** The activated JAKs interact with latent cytoplasmic proteins termed STATs and activate them through phosphorylation. **(3)** Subsequently, the activated STATs dimerise and **(4)** translocate into the nucleus where they induce transcription of certain genes by binding to specific sequences.

1.4 Aims of this thesis

Matrix Attachment Regions (MARs) are genomic sequences that mediate the anchoring of chromatin into loop domains. Although there is evidence implicating MARs in different aspects of genomic function, comprehensive knowledge of their function is still lacking since small-scale methods have limited their identification to very few regions of the genome. Are MARs characterised by a unique sequence consensus motif? How are they distributed across the genome? Are they static or dynamic elements of the genome? What role do they play in transcriptional regulation?

I have approached these questions by examining the tethering of a large locus, the human Major Histocompatibility Complex (MHC) in cell types exhibiting different patterns of MHC expression. Thus, the overall aim of this thesis is to improve our understanding of the relationship between chromatin anchors, or MARs and gene expression in the MHC. To this end:

1. The utility of a novel MHC tiling path array for mapping MARs was determined – Chapter 3.

2. MARs were mapped and characterised across the MHC locus in human fibroblasts – Chapter 4.

3. MARs were mapped and characterised across the MHC locus in IFN-γ treated human fibroblasts and B lymphoblastoid cells – Chapter 5.

2 Materials and Methods

2.1 Cell culture

Cells were grown on 75 cm^2 plastic culture flasks in incubators maintaining a temperature of 37 °C and a humidified atmosphere supplemented with 5% CO_2. Cell lines, growth media and other cell culture reagents were supplied by Cancer Research UK Cell Services (Clare Hall).

MRC5 cells are human lung fibroblasts derived from a male embryo (ECACC 97112601). They were cultured as monolayers in RPMI-1640 media supplemented with 10% foetal calf serum (FCS), 2mM L-glutamine, 100 units/ml penicillin V and 0.1 mg/ml streptomycin. MRC5 cells were passaged when they reached confluence, usually every 3-4 days. They were washed with phosphate-buffered saline (PBS) and detached by incubating them for 5 min with 3 ml of a solution containing porcine trypsin and EDTA per culture flask. The trypsin digestion was then blocked by the addition of an excess culture medium containing FCS. Cells were reseeded in two or three flasks at 30-50% confluence. Cell-surface expression of HLA molecules in MRC5 cells was induced by adding 200 IU/ml of human IFN-γ (recombinant Human IFN-γ R&D Systems) to the culture medium for 24 h (Darley et al., 1993).

HT1080 is a human fibrosarcoma cell line derived from a female (ECACC 85111505). HT1080 cells were cultured as monolayers in Dulbecco's Modified Eagle's Medium (DMEM) and supplemented with 5% FCS, 2mM L-glutamine, 100 units/ml penicillin V and 0.1 mg/ml streptomycin. They were usually passaged every 2-3 days as described for MRC5 cells.

PGF is a human EBV-transformed B lymphoblastoid cell line derived from a consanguineous male and is homozygous for the MHC (ECACC 9405342). PGF cells were grown in suspension in RPMI-1640 supplemented as described for MRC5 cells, and maintained at exponential growth by addition of fresh media every 2-3 days so that their concentration would not exceed 10^6 cells/ml.

Raji is a B lymphoblastoid cell line established from a Burkitt's lymphoma of a human male (ECACC 85011429). RJ2.2.5 is a CIITA-deficient mutant derived from the human Raji cell line (ICLC HTL02002). Both cell lines were grown in suspension using the same conditions described for PGF cells by the Immunopathology Laboratory at Geneva University.

2.2 Isolation of matrix attachment region and halo DNA fractions

The optimal time for nuclear protein extraction was resolved as described previously (Krawetz et al., 2005). PGF and MRC5 cells +/- IFN-γ were washed in PBS supplemented with 1 mg/ml bovine serum albumin (BSA) and Complete Protease Inhibitor (CPI) (Roche). The cells were re-suspended in PBS supplemented with CPI and lysed on ice for 15 min following addition of Nuclear Buffer (10mM Tris, pH 7.7, 100mM NaCl, 300mM sucrose, 3mM $MgCl_2$, 0.5% Triton-X 100 and CPI). The concentration of nuclei was adjusted to 1.5 x 10^5 /ml with PBS supplemented with CPI and 1.25 x 10^4 nuclei were attached to slides using a Shandon Cytospin 2 cytofuge (Thermo Shandon).

Halo Buffer (10mM Tris, pH 7.7, 10mM EDTA, 2M NaCl, 1mM dithiothreitol and CPI) was placed on the nuclei and incubated for varying times ranging from 0 to 15 min. The reaction was stopped by washing the slide for 1 min in a graded series of PBS washing solutions, and then the cells were dehydrated using a series of graded ice-cold ethanol washes. Each slide was then baked for 2 h at 72 °C. The nuclei were stained with DAPI and viewed using a Zeiss Axiophot microscope, equipped with a CCD-camera (Photometrics), which was used to obtain images.

For each cell type, the size of the halo structures at each time point was determined by measuring the average distance between the perimeter of the halo and the edge of the nuclear matrix of 30 cells in 3 biologically independent experiments. The minimum time required to reach maximum average distance between the edge of the nuclear matrix and the perimeter of the extracted DNA was considered optimal.

The isolation of MAR and halo DNA fractions was repeated five times for each cell line to obtain a total of five biological replicates. Two of these replicates were used for array hybridisations, and the other three for real-time PCR.

PGF and MRC5 cells +/- IFN-γ were washed in PBS supplemented with 1 mg/ml BSA and CPI. The cells were then re-suspended in 125 μl of PBS and 125 μl of 2x Nuclear Buffer and lysed on ice for 15 min. The nuclei were collected by centrifugation and incubated in 250 μl of 2x Halo Buffer on ice. The incubation time with 2x Halo Buffer was 12 min for nuclei from MRC5 cells +/- IFN-γ, while it was 9 min for nuclei from PGF. The reaction was then stopped with the addition of 9.5 ml Restriction Buffer (50 mM Tris, pH 8.0 and 10 mM $MgCl_2$) supplemented with 100 U of EcoRI and HindIII (Sigma). The DNA was digested at 37 °C for 4 h.

Following centrifugation at 16,000 g, the supernatant containing the halo fraction and the pellet containing the MAR-enriched fraction were placed in separate tubes. The MAR-enriched sample was washed three times with Restriction Buffer. Both samples were digested with 120 μg of proteinase K (Roche) in Protease K Buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 25 mM EDTA and 0.5% w/v SDS) overnight at 55 °C.

The DNA of both fractions was purified using a commercial purification kit (QIAmp DNA blood kit) and their concentration determined by using a NanoDrop ND-1000 which can quantify DNA in 1 ul samples without performing dilutions or washing cuvettes.

2.3 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was carried out using HT1080 cells following the method described in (Christova and Oelgeschlager, 2002) with minor amendments. Approximately 5 x 10^6 cells were cross-linked with 1% v/v formaldehyde at room temperature for 15 min under constant agitation. To stop the cross-linking reaction, glycine was added to a final concentration of 125 mM.

Cells were washed twice in ice-cold PBS and re-suspended in Lysis Buffer (50 mM HEPES, pH 7.9, 10 mM EDTA, pH 8.0 and 1% w/v SDS) supplemented with CPI.

They were then sonicated on ice using a sonicator mounted with a 2 mm diameter probe (Sonics) until cross-linked chromatin was sheared to an average DNA fragment length of 1−2 kb. The length of the DNA fragments was monitored by agarose gel electrophoresis.

After removing cell debris by centrifugation, the soluble cross-linked chromatin was diluted 1:10 in Immunoprecipitation (IP) Buffer (10 mM HEPES, pH 7.9, 1% Triton X-100 and 150 mM NaCl) supplemented with CPI, divided into aliquots and stored at -80 °C. Protein A−Sepharose CL4B (Pharmacia) was blocked with 1 mg ml^{-1} BSA (Roche) and 1 mg ml^{-1} herring-sperm DNA (Sigma) in IP Buffer for 4−6 h at 4 °C and washed three times with IP Buffer before use.

In order to pre-clear chromatin, blocked Protein A−Sepharose (50 μl of 50% slurry per ml of chromatin) was added to the soluble chromatin preparations for 2 h at 4 °C. The Protein A−Sepharose was then removed by centrifugation and the pre-cleared chromatin was incubated with antibodies against acetyl-histone H3-K9/K14 (Upstate, catalogue number 06-599) for 12−14 h at 4 °C. The immunoprecipitates were incubated with fresh blocked Protein A-Sepharose for 2 h at 4 °C and recovered by centrifugation.

The pellets were washed three times with IP Buffer, three times with Washing Buffer (10 mM Tris, pH 8.0, 0.25 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate and 1 mM EDTA, pH 8.0) and three times with Tris-EDTA (TE), pH 8.0. The pellets were re-suspended in TE, pH 8.0 and digested with 50 μg ml^{-1} RNAse A (Roche) for 30 min at 37 °C. The immunoprecipitated DNA was mixed with 250 μg ml^{-1} proteinase K (Roche), SDS to 0.25% w/v and incubated for 12 h at 37 °C. The formaldehyde cross-links were then reversed by incubating the chromatin preparations for 6 h at 65 °C. The chromatin was purified using a commercial purification kit (QIAmp DNA blood kit) and the DNA concentration determined by using a NanoDrop ND-1000.

2.4 *Construction of a MHC tiling path array*

The tiling path array of the MHC was constructed and optimised in collaboration with the Immunogenomics Group at the Wellcome Trust Sanger Institute (Tomazou et al., 2008). A total of 1747 overlapping plasmid clones covering the MHC region were used

to generate the array. Of those, 1662 clones (average insert size of 2 kb) were picked from the HapMap chromosome 6 library at the Wellcome Trust Sanger Institute. The remaining 85 clones were generated by cloning gap-spanning PCR amplicons (average insert size 332 bp). Across the MHC, about 12 kb of repeat-rich regions were refractory to PCR amplification and could not be included in the array. Therefore, the overlapping plasmid clones cover 99.67% of the entire MHC region.

Vector-specific PCR was performed using each clone to produce amino-linked amplicons. The PCR reactions were carried out in 50 mM KCl, 5 mM Tris, pH 8.5 and 2.5 mM $MgCl_2$ (10 min at 95 °C followed by 35 cycles of 95 °C for 1 min, 60 °C for 1.5 min, 72 °C for 7 min and a final extension of 72 °C for 10 min) using the forward primer 5'-CCCAGTCACGACGTTGTAAAACG-3' and the reverse primer 5' AGCGGATAACAATTTCACACAGG-3'. In order to generate strand-specific array probes, two separate PCR reactions were performed for each clone, in one case using a 5'-amino-linked primer for the forward strand, and in the other case, for the reverse strand.

After assessing the products by agarose gel electrophoresis, they were re-suspended in Spotting Buffer (250 mM sodium phosphate, pH 8.5, 0.00025% w/v sarkosyl and 0.1% sodium azide) and filtered (Multiscreen-GV filter plates, Millipore). The products were spotted onto amine binding slides (CodeLink, GE Healthcare) at 20–25 °C, 40–50% relative humidity. After incubating the slides overnight in a humid chamber, they were treated with 1% ammonium hydroxide for 5 min and then 0.1% SDS w/v for 5 min. They were then denatured with 95 °C ddH_2O for 2 min, rinsed in ddH_2O and dried by centrifugation.

Each clone was printed four times across the array to minimize spatial effects. Test sequencing was carried out on 15% of the amplified PCR products. The results show that the sequences of ~97% of the selected PCR products matched the expected regions of the MHC. From this partial analysis, it is possible to extrapolate that ~3% of the spots of the MHC tiling path array do not represent the expected MHC sequences.

2.5 Random labelling of DNA samples

DNA samples were labelled with Cy3- and Cy5-labelled dCTPs (Amersham Biosciences) using a random prime labelling system (Bioprime DNA Labelling System; Invitrogen) as described previously (Mulholland et al., 2006).

A pre-mixture was prepared containing between 0.3 and 0.5 μg of DNA, 60 μl of 2.5x Random Primer Mix. The volume was made up to 126 μl with distilled water. After denaturing the DNA template for 10 min at 100 °C, the pre-mixture was snap-cooled on ice. A solution containing 15 μl of 10x dCTP mix, 6 μl of Cy3- or Cy5-labelled dCTP (Perkin Elmer) and 3 μl of Klenow fragment was added to the pre-mixture while on ice and the sample was incubated at 37 °C overnight. The reaction was stopped by adding 15 μl of 0.5 M EDTA.

The labelled samples were purified on Micro-spin G50 columns (Amersham) to remove unincorporated nucleotides. Fluorescent dye incorporation was monitored using the Nanodrop ND-1000 spectrophotometer.

2.6 Array hybridisation and washes

The Cy3- and Cy5-labelled DNA samples were hybridised onto the MHC tiling path array using a procedure previously described in (Fiegler et al., 2003). In one tube, equal amounts of Cy3- and Cy5-labelled probe were combined with 100 μg human Cot-1 DNA (Roche) and precipitated with absolute ethanol precipitation using the standard methodology. In a second tube, 300 μg of herring sperm DNA (Sigma) was added to 50 μg of human Cot-1 DNA and precipitated in a similar manner.

The pellets of the first and second tubes were re-suspended in 60 and 200 μl, respectively, of Hybridisation Buffer (50% v/v deionised formamide, 10% w/v dextran sulphate, 0.1% v/v Tween 20, 2x SSC, and 10 mM Tris, pH 7.5). In order to re-suspend the pellets uniformly, after applying the buffer to each tube, the samples were kept at 80 ^{o}C on a heat block for 10 min with vortexing every 2 min.

The content of the second tube was applied to the MHC tiling path array, which was placed in a humid chamber saturated with 20% v/v formamide and 2x SSC. The chamber was sealed air-tight and placed in a hybridisation oven with gentle rocking for 1 h at 37 °C. Meanwhile, 40 ng of yeast tRNA (Sigma) was added to the first tube, which was then incubated at 37 °C for pre-annealing. Subsequent to array pre-hybridisation the contents of the first tube was applied to the array and placed in the same humid chamber and allowed to hybridise for 48 h at the same conditions as described above.

The MHC tiling path array was removed from the chamber and placed in a Coplin jar containing 2x SSC and 0.03% w/v SDS to wash off any excess hybridisation solution. The slide was then transferred to another Coplin jar and washed in 2x SSC and 0.03% w/v SDS for 5 min at 65 °C, three times in 0.2x SSC for 20 min at room temperature and in PBS and 0.05% v/v Tween 20 for 10 min at room temperature. After rinsing in distilled water, the slide was spin dried at 1000 rpm for 5 min using a swinging rotor and then placed in a dark chamber at room temperature.

2.7 Array scanning and data analysis

Arrays were scanned at 532 nm (Cy3) and 635 nm (Cy5) using the confocal scanner ScanArray 4000 (GSI Lumonics). For each clone, a ratio of Cy5 to Cy3 fluorescent intensity was calculated by processing the array images in TIGR Spotfinder (http://www.tm4.org/spotfinder.html). The median spot intensities were subtracted from the median intensities of each spot for each Cy5 and Cy3 images.

Analysis of the data was carried out with the contribution of Richard Mitter, at the Bioinformatics and Biostatistics group, London Research Institute Cancer Research UK, using the open-source software Bioconductor (Smyth, 2004).

The Cy5 to Cy3 $\log_2$ ratios were normalized by subtracting the weighted median for each array. Intensities were normalised between arrays using the "quantile" method of "NormaliseBetweenArrays" function from the Bioconductor package *limma* to ensure that the intensities have the same empirical distribution across arrays and across channels. For each clone, a mean $\log_2$ was calculated using the four replicates within

each array. Pearson Correlation Coefficients were calculated between all arrays and found to be strongest between duplicates. Summarised $\log_2$ ratios were calculated by taking a mean across duplicate arrays.

Data points derived from clones constituted entirely of repetitive sequences were excluded from the analysis (Appendix I). Repetitive sequences were defined as in the 'all repeats' track in the Ensembl browser (Hubbard et al., 2007). Dye swap hybridisation for each pair of arrays was performed to normalise for any dye bias. The genomic position of the BACs used to generate the MHC tiling path arrays was determined using the NCBI build 34 assembly of the human genome.

2.8 Real-Time PCR

SYBR Green real-time PCR was used to analyse the process of DNA amplification quantitatively throughout the PCR reaction. SYBR is a cyanine dye that binds to the minor groove of double-stranded DNA and is fluorescent when bound.

PCR primers were designed using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Real-time PCR was performed using SYBR Green Jumpstart Taq Ready Mix kit (Sigma Aldrich, Missouri, USA). PCR reactions were prepared in 50 μl reaction volume and consisted of 25 μl of SYBR Green Jumpstart Taq Ready Mix, 100 ng of DNA template and forward and reverse primer mix at final concentration of 500 nM.

The PCR cycling conditions consist of an initial denaturation for 5 min at 95 °C, 40 cycles of denaturation at 95 °C for 40 sec, annealing at 55 °C for 30 sec and extension at 72 °C for 30 sec, followed by a final 5 min extension step at 72 °C. After completion of the amplification, a melting curve of each product was obtained by heating the samples from 72 °C to 95 °C and monitoring fluorescence every 1 °C increase in temperature to determine the specificity of the PCR reaction. These reactions were performed using the Opticon2 (MJ Research) thermal cycler.

In order to quantify the target regions in the DNA samples, standard curves were produced for each primer pair from PCR reactions with known amounts of genomic

DNA. Data analysis was performed using different versions of the Opticon (MJ Research) amplification detection software.

2.9 Threshold of array enrichment identifying matrix attachment regions

In order to identify MARs based on the array data, a threshold of array enrichment was determined by comparing the array enrichment values with the quantitative real-time PCR enrichment of selected MHC clones using MAR and halo DNA fractions.

The selected MHC clones were characterised by MAR/halo array enrichments ranging from approximately -1 to 1 ($\log_2$). Each real-time PCR reaction was carried out as described in section 2.9. The experiments were performed three times using MAR and halo DNA fractions from three independent populations of untreated MRC5 cells (samples 3, 4 and 5).

The data show that the clones having array enrichment values greater than 0.4 ($\log_2$), are consistently confirmed to be MARs (Table 2.1). Therefore, genomic regions with array enrichment greater than 0.4 were thus designated as MARs.

MHC region	MAR/Halo predicted by array enrichment	Array enrichment value MAR/halo (log_2)	RT-PCR enrichment MAR/halo (log_2)
30455483-30457800	MAR	0.999	3.234
31955316-31957399	MAR	0.803	2.526
32591170-32594283	MAR	0.597	3.624
29933548-29935949	MAR	0.406	1.902
29739385-29741249	MAR	0.201	-0.919
30118308-30120829	MAR	0.008	-1.485
33051852-33053947	Halo	-0.202	-0.823
32127610-32130779	Halo	-0.400	-2.724
33190199-33192634	Halo	-0.602	-1.251
31236895-31239353	Halo	-0.811	-1.923
31649837-31652271	Halo	-1.005	-2.014

Table 2.1 Comparison between real-time PCR (RT-PCR) data and array enrichment values.

2.10 Expression Array

The gene expression of PGF and MRC5 cells +/- IFN-γ was determined using Affymetrix U133 plus 2.0 GeneChips (Affymetrix, Santa Clara, CA). Total RNA was extracted using the Qiagen RNeasy Mini Kit following the manufacturer's protocol. Briefly, 5 x 10^6 cells were lysed and homogenised in the presence of a highly denaturing guanidine-thiocyanate–containing buffer, which inactivates RNases to ensure purification of intact RNA. 70% ethanol was then added to the lysate to promote selective binding of RNA to the membranes of the RNeasy Mini spin columns. A series of washes with RNeasy washing buffers was then performed to remove contaminants before eluting the total RNA from the RNeasy mini columns with 60 μl of RNase-free water. The bind, wash and elution steps were performed by centrifugation. The amount of RNA was measured using the Nanodrop ND-1000 spectrophotometer and the RNA quality assessed using the Agilent 2100 Bioanalyser.

The samples were prepared and hybridised to the microarrays according to the Affymetrix Technical Manual (www.affymetrix.com/support/technical/manual/

expression_manual.affx). Briefly, total RNA (10 μg) was converted into double-stranded cDNA using the reverse transcriptase and oligo-dT primers of the One-Cycle cDNA Synthesis Kit. *In vitro* transcription of biotin-labelled antisense RNA (cRNA) was carried out using the cDNA as a template. Prior to hybridisation, the cRNA was fragmented in Affymetrix Fragmentation Buffer at 94 °C for 35 min to produce 25-200 bp fragments. Approximately 15 μg of labelled, fragmented cRNA was added to a hybridisation cocktail, which was then hybridised to U133 Plus 2.0 GeneChips for 16 h at 45 °C with gentle rocking. GeneChips were washed with a series of non-stringent and stringent solutions containing variable amounts of morpholino ethane sulfonic acid, Tween 20, sodium chloride, sodium phosphate and EDTA. The microarrays were then stained with streptavidin phycoerythrin, and the fluorescent signal was amplified using a biotinylated antibody solution. Fluorescent images were detected using a GeneChip Scanner 3000, and expression data extracted using the GeneChip Operating System v 1.1 (Affymetrix).

Statistical analysis was performed using the open source statistical environment R (www.r-project.org) and Bioconductor packages (Smyth, 2004). Probe set expression measures were calculated using the "Affy" package's Robust Multi-chip Average (RMA) default method. Probe sets exhibiting a >2 fold change were called differentially expressed. Where multiple probe sets mapped to the same gene, a mean fold change was taken. All expression data were deposited in NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE10880.

2.11 FISH and image analysis

FISH was performed on PGF and MRC5 cells +/- IFN-γ to examine the higher-order chromatin architecture of the MHC. The PAC RP1-172K2 probe was selected for the MHC class II gene *HLA-DRA* and obtained from the Wellcome Trust Sanger Institute, UK. Probe DNA was labelled with digoxigenin-11-dUTP (Roche) by nick-translation using the enzyme mix from BioNick kit (Invitrogen) and a digoxigenin labeling mixture replacing biotin 14-dATP (Roche Diagnostics). A solution containing 5 μl of digoxigenin labelling mixture, 5 μl of 10x enzyme mixture, 1 μg of probe DNA and made up to 40 μl with water was incubated at 16 °C for approximately 85 min to

produce 500 bp fragments. The size of the fragments was examined by agarose gel electrophoresis. In order to inactivate the enzymes, 5 μl of EDTA was then added and the mixture heated to 60 °C for 5 min.

PGF and MRC5 cells +/- IFN-γ were fixed in a methanol acetic acid (3:1) solution and placed on glass microscope slides. Prior to hybridisation, the chromosomes on each slide were aged with microwave irradiation for 2.5 min using a microwave (Sharp Carousel II) set at medium/high power. To denature the chromosomes, 90 μl of denaturation mixture (70% v/v formamide, 2x SSC) was placed on the slide under a 22 x 50 mm coverslip and the slide heated on a hot plate at 73 °C for 3 min. The slide was then washed with 2x SSC and dehydrated in alcohol.

Labelled DNA (300 ng) was precipitated with human Cot-1 DNA, salmon sperm DNA, NaAc and absolute alcohol and then re-suspended in hybridisation mixture (50% formamide, 2x SSC and 10% w/v dextrane sulphate). The probe was denatured on a heating block at 85 °C for 5 min and incubated at 37 °C for 30 min. The slide was then pre-warmed on a heating block at 37 °C, the denatured probe applied to the slide with FITC-labelled chromosome 6 paint (Cambio) and the slide placed in a moist chamber at 37 °C overnight.

The slide was washed three times in 50% v/v formamide, 2x SSC at 42 °C and then three times in 1x SSC at 60 °C. The hybridisation area was then blocked in 90 μl of blocking solution (4x SSC, 0.1% v/v Tween 20 and 5% w/v skimmed milk powder (Marvel)) and incubated at 37 °C for 25 min. The blocking solution was drained from the slide and the detection antibody applied and covered with a coverslip (22 x 50 mm).

Anti-avidin-FITC 1:300 and anti-digoxigenin-rhodamine 1:100 antibodies were used for detection by incubation at 37 °C for 45 min followed by three washes with 4x SSC and 0.1% v/v Tween 20 at 42 °C. The chromosomes were then counterstained with DAPI (200 ng/ml) in Citifluor antifade solution and stored at 4 °C in the dark. Slides were examined with a Zeiss Axiophot epifluorescence microscope.

Separate grey-scale images were recorded with a cooled CCD-camera (Hamamatsu). They were then pseudocoloured and merged using SmartCapture X software (Digital Scientific, Cambridge). The position of signal from the *HLA-DRA* probe was analysed

in relation to the domain defined by the chromosome 6-specific paint. Each chromosome homologue was counted separately and a score derived for the percentage of loci that were located on an external chromatin loop. Experiments were repeated twice on cells from independent cultures.

3 Validation of a tiling path array for the MHC

3.1 Introduction

The MHC is one of the most important genomic regions for protection of the organism against diseases. While much effort has been made to understand the genetics of this locus, contributing greatly to our knowledge on the regulation of MHC genes, further studies are necessary to elucidate the relationship between the three-dimensional properties of the MHC and its function. This region constitutes an ideal model to investigate chromatin organisation as it contains genes that are constitutively expressed, tissue-specific or inducible (Boehm et al., 1997; Rohn et al., 1996). Furthermore the MHC has distinct structural configurations dependent on its transcriptional status (Christova et al., 2007; Volpi et al., 2000).

The development of assays using DNA microarrays and high-throughput sequencing and their application in combination with other techniques, including ChIP, has allowed the study of chromatin structure on a genome-wide scale. These technologies are applicable to properties of chromatin such as nucleosome positioning, DNA methylation, chromatin-modifying enzymes, histone and non-histone structural proteins (Schones and Zhao, 2008). In addition, high-throughput approaches can be used to study higher-order chromatin architecture. For instance, the chromosome conformation capture approach allows the identification of multiple long-range DNA interactions (Simonis et al., 2006; Zhao et al., 2006). Analysis of cells at different stages of differentiation or exhibiting different transcription and replication profiles enables the role of chromosome architecture in genomic function to be determined.

In order to facilitate the comprehensive study of MHC architecture, I generated a novel tiling path array covering the locus, in collaboration with the Immunogenomics Group at the Sanger Institute. The array was constructed to determine the DNA-protein interactions that contribute to the organisation of the MHC and how the structural properties of the MHC affect its function. It was also designed to be compatible with

various assays including ChIP, methylated DNA immunoprecipitation (MeDIP) and array comparative genomic hybridisation (aCGH).
The array platform was constructed utilising a public clone resource (http://www.hapmap.org) at a fraction of the costs associated with commercially available arrays. A study by Tomazou et al. (2008) showed that the MHC array can be successfully used for the identification of tissue-specific differentially methylated regions (Tomazou et al., 2008). Here, the identification of the Class II Transactivator (CIITA) binding sites and the analysis of histone H3-K9/K14 acetylation across the MHC were used to validate the array since they are both well described (Krawczyk et al., 2004; Krawczyk et al., 2008; Liang et al., 2004; Roh et al., 2005).

CIITA induces transcription of the classical and non-classical class II genes by interacting with transcription factors at the conserved proximal-promoter SXY motif and the distal enhancer-module S'-Y'(Gobin et al., 1997; Harton and Ting, 2000; Krawczyk et al., 2004; Martin et al., 1997). Acetylation of the amino terminal lysines of histones H3 is associated with induction of transcription and is thought to disrupt condensed chromatin to allow accessibility of transcription factors and RNA polymerase (Beresford and Boss, 2001; Grunstein, 1997; Utley et al., 1998). The histone H3-K9/K14 acetylation profile of the MHC was tested as it is a feature of 5' regions of transcriptionally active human genes (Liang et al., 2004; Roh et al., 2005).

The validation experiments were performed using the MHC tiling path array in combination with ChIP. The data show that the arrays identify previously described CIITA-binding sites with high reliability as well as novel putative CIITA-binding sites, which are mostly located in the promoter and regulatory region of MHC genes. The histone study confirms that the highest levels of H3-K9/K14 acetylation are at the promoter regions of constitutively expressed genes. Taken together these experiments demonstrate the utility of the arrays for the profiling of DNA-protein interactions and the epigenetic status of the region.

3.2 Results

3.2.1 MHC tiling path arrays

A tiling array, encompassing almost the entire 4 Mb MHC region (99.67%) at a resolution of 2 kb, was constructed. The array contains 7832 probes consisting of two replicates of both forward and reverse strands of MHC sequences and appropriate controls. Approximately 97% of these probes were estimated to be informative based on the quality controls described in Chapter 2.

3.2.2 Analysis of technical variation

Regardless of the experiment performed, the identification of MHC regions that are truly enriched for the factor of interest is of critical importance. To distinguish between the real enrichment and the background enrichment, which occurs as consequence of technical noise, the experimental variation was calculated using a series of self-self hybridisations.

The DNA samples used consisted of genomic DNA from human MRC5 fibroblast cells sonicated to an average size of 1 kb. These samples were differentially labelled with dCTP-Cy3 and dCTP-Cy5 fluorophores and hybridised to the MHC tiling path array. The slides were then washed and scanned and acquired images were analysed with an imaging software (Spotfinder). This experiment was carried out three times. The Cy3/Cy5 ratio of each probe on the array was used as an indication of the representation of region covered by the probe in the two samples. The ratios of the fluorescent dyes of each locus were $\log_2$ transformed, and the combined means calculated and subsequently plotted on the UCSC human genome database (http://genome.ucsc.edu, May 2004 Assembly) (Fig 3.1). Array probes consisting entirely of repetitive sequences, as listed in the previous chapter, were excluded from the analyses.

Findings are shown as $\log_2$ ratios, with a ratio of zero representing no enrichment of either sample. As samples from the same genomic DNA were co-hybridised to the MHC tiling path array, the $\log_2$ ratios were expected to remain close to 0 and any deviations to be attributable to experimental errors. The average $\log_2$ ratio was 0.01. The graphical representation of the results shows that there are no significant enrichments across the MHC locus. The experimental variation was determined by calculating the standard deviation of the $\log_2$ ratios of the array probes for each experiment. The average standard deviation was as low as 0.11.

In order to validate the array results, eight regions across the MHC were randomly selected and subjected to independent real-time PCR analysis. The $\log_2$ array enrichment of the selected MHC regions varied from –0.04 to. 0.10. For all 8 regions, the amount of DNA was found to be almost equivalent in both genomic samples. This is consistent with the array data suggesting that the MHC tiling path array is suitable for the identification of differential DNA enrichment (Fig 3.2).

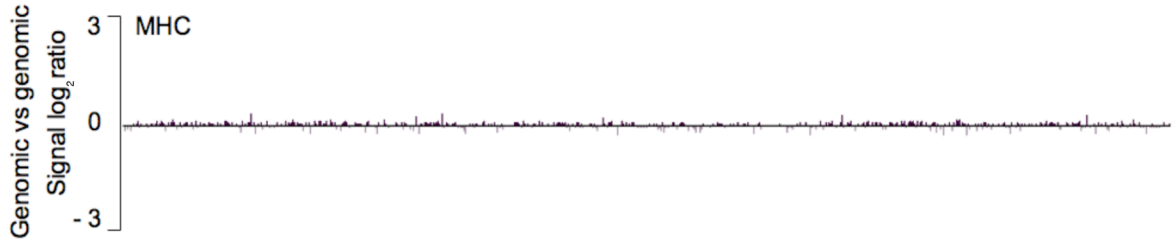

Figure 3.1 MHC plot of self-self experiments.

The plot was obtained by co-hybridising the two samples of genomic DNA from MRC5 cells labelled with Cy3 and Cy5 fluorophores onto the MHC tiling path array. The experiment was carried out three times. Each bar represents a clone from the MHC locus. The height of each bar shows the average $\log_2$ enrichment level of each sample at a given MHC region. Positive ratios represent enrichments of a given MHC region in the Cy5-labelled samples while negative regions represent enrichments of a given MHC region in the Cy3-labelled samples.

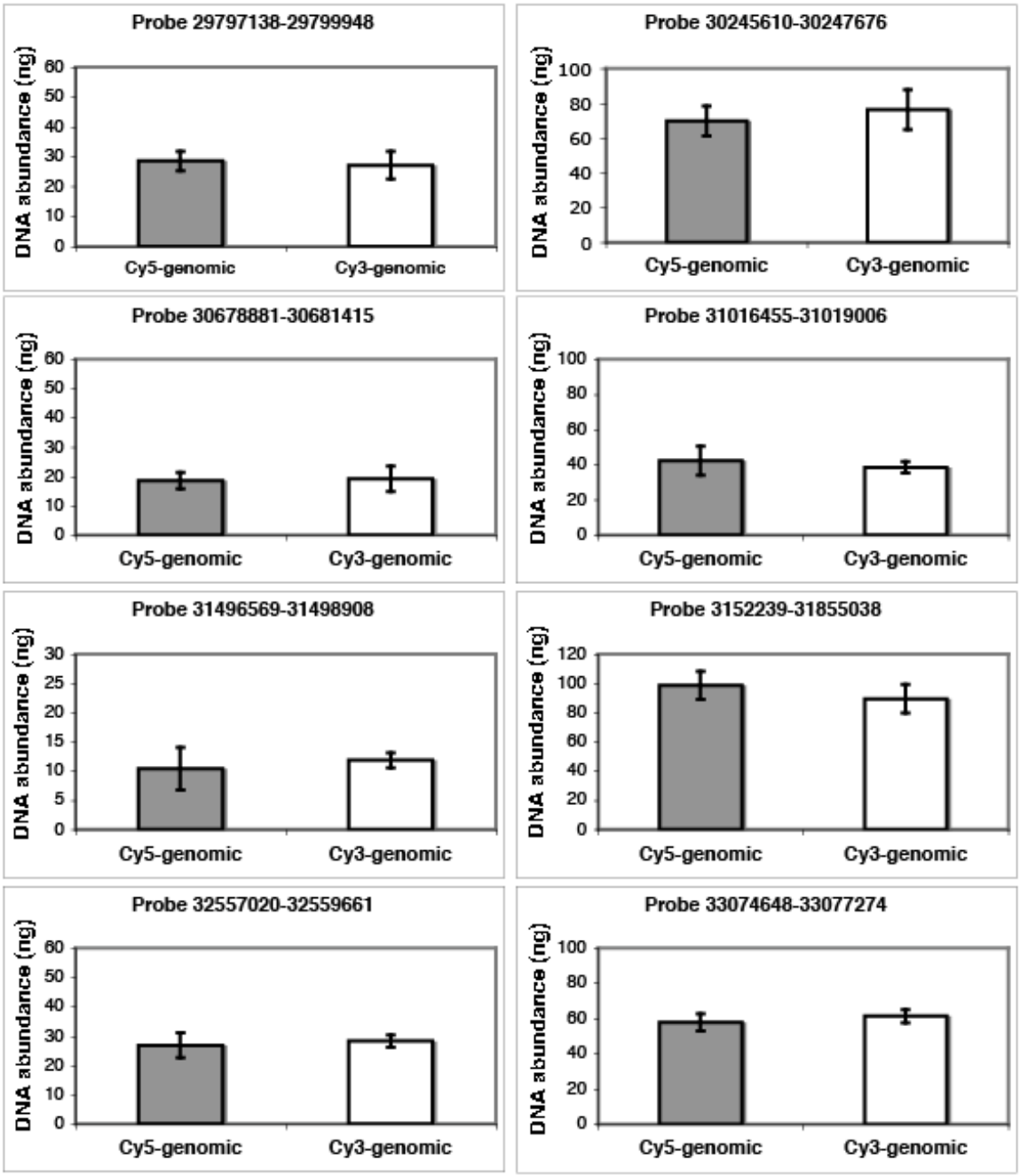

Figure 3.2 Quantitative real-time PCR analysis of eight regions across the MHC locus.

Real-time PCR was performed using two samples of genomic DNA from MRC5 cells and primers spanning eight MHC regions exhibiting no significant difference in enrichment of Cy5- and Cy3-labelled genomic DNA in the array experiments. Real-time PCR indicated that the quantity of those amplicons is almost equivalent in both Cy5- and Cy3-labelled genomic DNA samples, thus validating the array data. Bars represent the mean of two experiments and show the standard deviation.

3.2.3 CIITA-binding sites across the MHC

ChIP was performed using polyclonal anti-CIITA in the human Burkitts' lymphoma cell-line Raji and a CIITA-deficient cell-line generated *in vitro* from Raji, RJ2.2.5. Raji cells express CIITA constitutively and the ChIP samples from these cells were enriched in CIITA target sequences. Conversely RJ2.2.5 cells do not synthesise CIITA and the resulting ChIP DNA sample was enriched only in non-specific immunoprecipitated DNA sequences. The ChIP DNA products from both cell lines were competitively hybridised to the MHC tiling path arrays. ChIP assays were carried out by the Immunopathology Laboratory at the Geneva University using the method described by Masternak et al. (2003) (Masternak et al., 2003).

Due to the small amounts of DNA obtained from this procedure, two rounds of ligation mediated PCR were carried out prior to competitive hybridisation. The ChIP samples from Raji and RJ2.2.5 cells were then labelled with Cy3 and Cy5 fluorophores, respectively, and co-hybridised to the MHC tiling path array. A dye swap experiment was carried out to control for the efficiency of the labelling. The ratios of the fluorescent dyes of each locus were $\log_2$ transformed, and the combined mean values calculated and plotted on the UCSC human genome database (http://genome.ucsc.edu, May 2004 Assembly). The graphical representation of the results shows that, although each class of the MHC contains CIITA-binding sites, the highest peaks representing CIITA enrichments are present across the MHC class II (Fig 3.3). The DNA regions enriched in the ChIP samples from RJ2.2.5 cells are expected to represent enrichment in immunoprecipitated DNA sequences derived from the nonspecific binding of the antibody and subsequent DNA amplification.

The best-documented targets of CIITA are the regions proximal to the classical and non-classical MHC class II genes, which contain the SXY and S'-Y' modules. These sequences were used as positive controls to validate the MHC tiling path arrays. A threshold of four standard deviations was used to select CIITA-enriched clones for further analysis. The application of this threshold identified 27 CIITA-enriched loci, of which 14 overlapped with the SXY and S'-Y' modules of the MHC class II genes (Table 3.1). No data were available for the S'-Y' module of *HLA-DRA* since it is a region refractory to PCR amplification due to repeats, thus, it could not be included in

the tiling path array. The S'-Y' module of *HLA-DQA1* is the only positive control represented in the MHC tiling path array, which is not confirmed to be highly enriched in CIITA.

Additional positive controls include the SXY and S'-Y' modules of MHC class I genes. The array results revealed that these modules are among the 27 highest CIITA-enriched sites across the MHC locus (Table 3.2). This list also includes sequences that were not candidate CIITA-binding sites. These regions were designated as novel putative CIITA-enriched sites (Table 3.3). As these sequences were more CIITA-enriched than positive controls, their enrichment is unlikely to be attributable to stochastic differences in the competitive hybridisation of targets to the probes on the arrays and technical variation. Of these novel putative CIITA-binding sites, four overlapped transcription initiation sites and two were located in the intergenic region downstream of the start site. Several of the novel putative CIITA-binding sites are either within or proximal to genes that contribute to a range of clinical conditions (Aquino-Galvez et al., 2008; Heinzelmann-Schwarz et al., 2004; Matsuzaka et al., 2002).

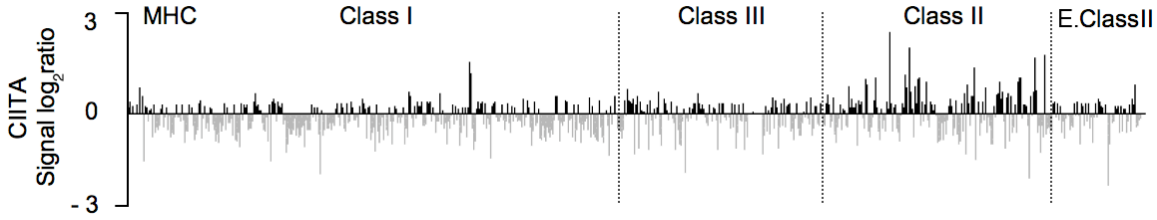

Figure 3.3 CIITA-binding sites across the MHC.

The plot was obtained co-hybridising Cy3- and Cy5-labelled ChIP DNA from Raji and RJ 2.2.5 to the MHC tiling path array. The values above 0 represent enrichment in ChIP DNA from Raji. The values below 0 represent enrichment in ChIP DNA from RJ 2.2.5. The experiment was carried out twice. The height of each bar shows the average log2 enrichment.

Gene	Description	Clone co-ordinates	Module
HLA-DPA1	MHC class II DP	33148418-33151260	SXY
HLA-DPB1	MHC class II DP	33150708-33153101	SXY
HLA-DRB1	MHC class II DR	32663263-32666061	SXY
HLA-DRB5	MHC class II DR	32605193-32606885	SXY
HLA-DRA	MHC class II DR	32513678-32516071	SXY
HLA-DQA1	MHC class II DQ	32712264-32714547	SXY
HLA-DQB1	MHC class II DQ	32740888-32743557	SXY
HLA-DQB2	MHC class II DQ	32838552-32841749	SXY
HLA-DOA	MHC class II DO	33084323-33086976	SXY
HLA-DOB	MHC class II DO	32890054-2892994	SXY
HLA-DMA	MHC class II DM	33027231-33028946	SXY
HLA-DMB	MHC class II DM	33015774-33018132	SXY
HLA-DRB1	MHC class II DR	32680962-32683760	S'-Y'
HLA-DMA	MHC class II DM	33042435-33045220	S'-Y'

Table 3.1 Positive controls within the MHC class II region identified by the MHC tiling path arrays.

Gene	Description	Clone co-ordinates	Module
HLA-A	MHC class I A	30016538-30019300	SXY
HLA-B	MHC class I B	31431405-31434315	SXY
HLA-C	MHC class I C	31347352-31349944	SXY
HLA-E	MHC class I E	30563893-30565706	SXY
HLA-F	MHC class I F	29797138-29799948	SXY
FLJ4522	Hypothetical protein LOC441140	30333825-30336492	S'-Y'
TRIM26	Tripartite motif-containing 26	30288220-30290560	S'-Y'

Table 3.2 Additional positive controls identified by MHC tiling path arrays.

Gene	Description	Clone co-ordinates	Location relative to gene start site (bp)
HLA-G	MHC class I G	29901199-29903435	0
DPCR1	Diffuse panbronchiolitis critical region 1	31024896-31027538	0
TRIM40	Tripartite motif-containing 40	30211416-30214316	0
TUBB	Tubulin-β	30798378-30800913	+2242
MICA	MHC class I chain-related gene A protein	31505209-31507825	+25859
TAP1	Transporter 1 ATP-binding cassette sub-family	32928521-32931366	0

Table 3.3 Novel putative CIITA-binding sites identified by ChIP-MHC tiling path arrays.

3.2.4 Quantitative real-time PCR of ChIP DNA from Raji and RJ2.2.5 cells

In order to confirm the array results, six DNA regions enriched in the ChIP sample from Raji cells and six DNA regions enriched in the ChIP sample from RJ2.2.5 cells were subjected to independent quantitative real-time PCR analysis. This validation was performed using the ChIP DNA samples used for the microarray experiments.

Each amplicon designed across the selected regions was quantified using standard curves produced for each primer pair from real-time PCR reactions with known amounts of genomic DNA. PCR reactions were performed at the same time using the same cycling conditions and with equal amounts of template and master mix in order to compare the quantity of each amplicon between the Raji and RJ2.2.5 ChIP samples.

The DNA regions enriched in the ChIP samples from Raji cells, representing sequences bound to CIITA, overlapped the SXY module of *HLA-DRA, HLA-DMA, HLA-F, HLA-C* and the transcription start site of *TAP1* and *HLA-G*. The DNA regions enriched in the ChIP samples from RJ2.2.5 cells, which instead represented enrichment in non-specifically immunoprecipitated DNA sequences, overlapped the intragenic regions of *MSH5, CFB, C4B, TNXB, COL11A2* and *CSNK2A1*. The real-time PCR results were consistent with the array data indicating that the MHC tiling path array is suitable for the study of DNA-protein interactions (Fig 3.4-3.5).

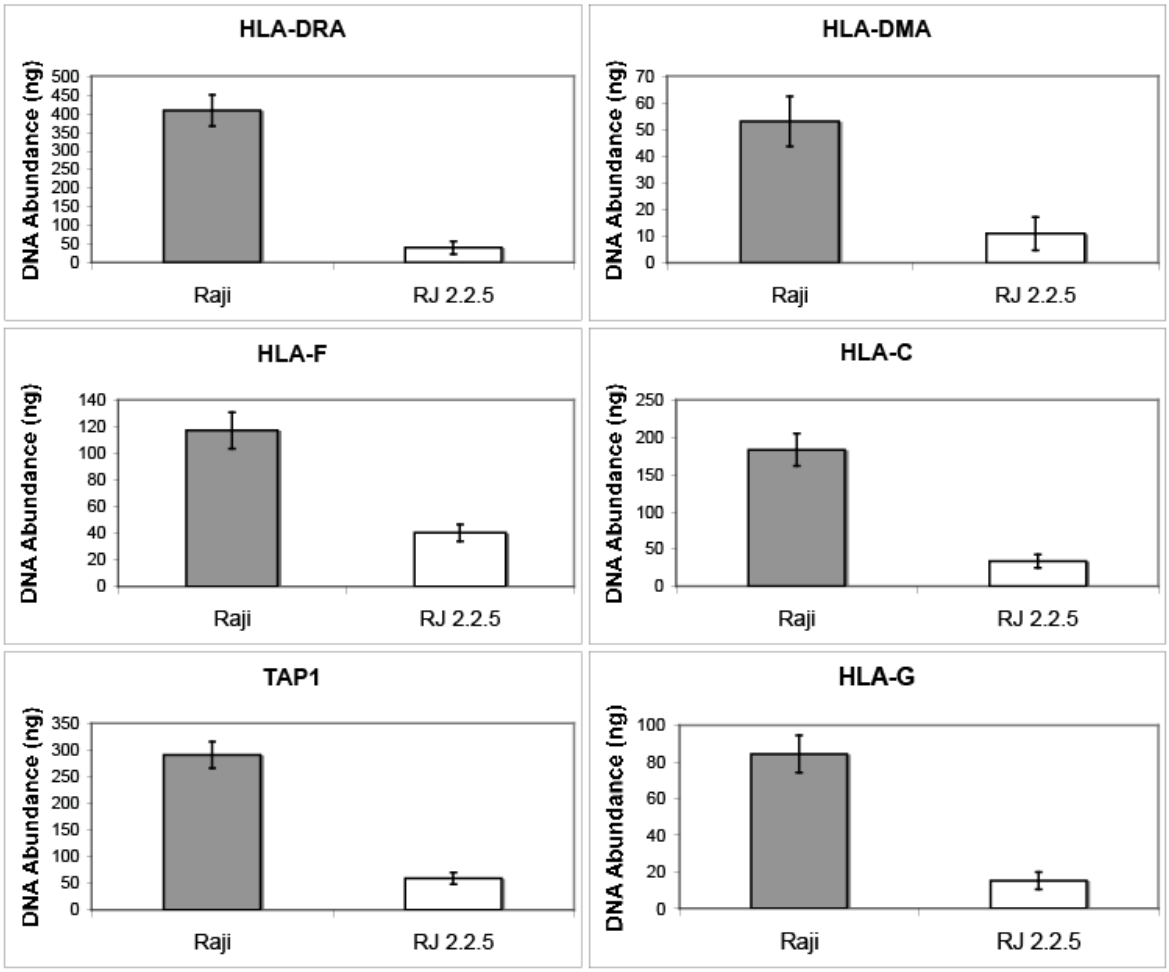

Figure 3.4 Quantitative real-time PCR analysis at CIITA-enriched loci in Raji and the derivative CIITA negative cell line RJ2.2.5.

Real-time PCR was carried out using two samples each of ChIP DNA from Raji and RJ 2.2.5 cells and primer pairs spanning regions of high CIITA enrichment according to the array data. The DNA regions overlapping the SXY modules of *HLA-DRA, HLA-DMA, HLA-F, HLA-C* and the transcription start sites of *TAP1* and *HLA-G* were found to be highly enriched in the Raji ChIP samples and thus in CIITA target sequences. Bars represent the mean of two experiments and show the standard deviation.

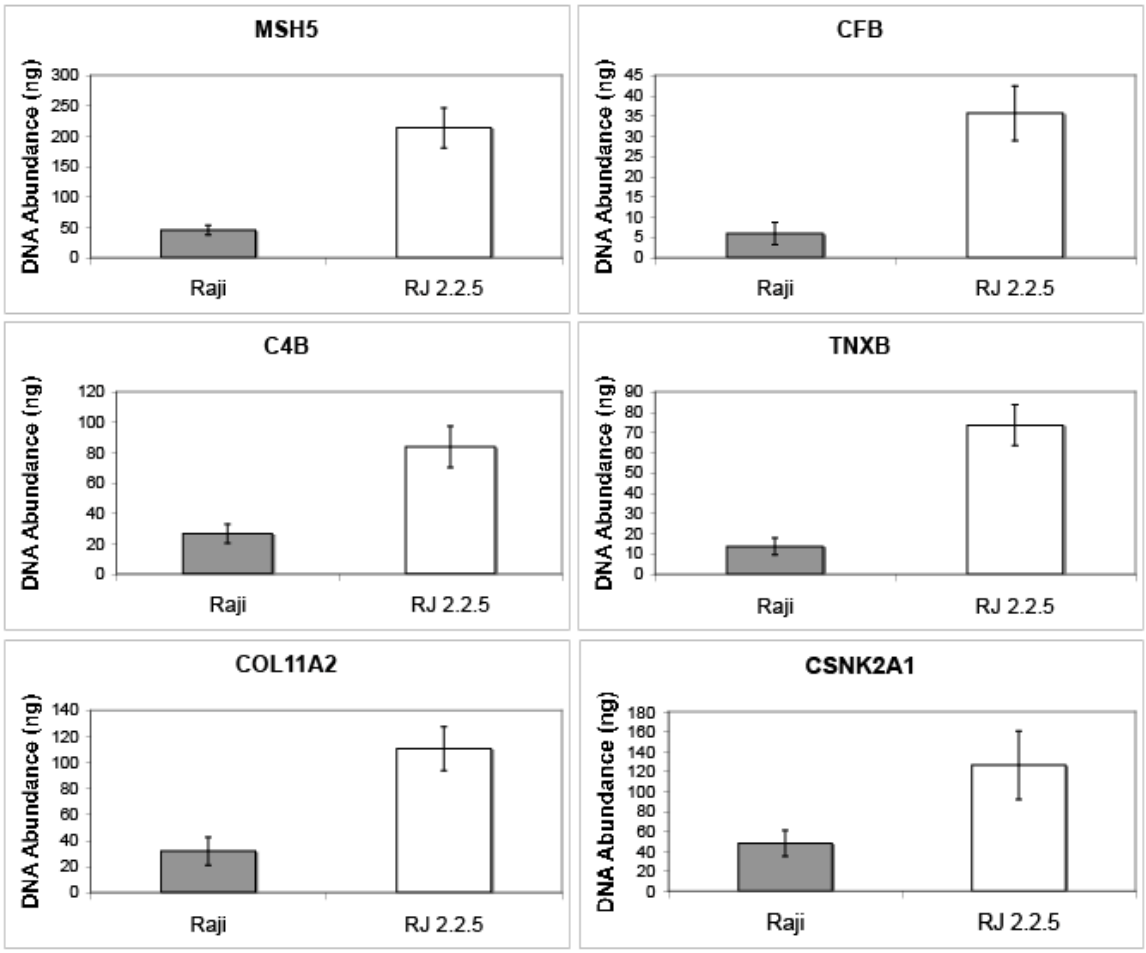

Figure 3.5 Quantitative real-time PCR analysis at loci enriched in non-specifically immunoprecipitated DNA sequences.

Real-time PCR was carried out using two samples each of ChIP DNA from Raji and RJ 2.2.5 cells and primer pairs within regions of no CIITA enrichment according to the array data. DNA sequences located in the intragenic region of *MSH5, CFB, C4B, TNXB, COL11A2* and *CSNK2A1* were found to be highly enriched in the RJ2.2.5 ChIP samples, which contain non-specifically immunoprecipitated DNA sequences. Bars represent the mean of two experiments and show the standard deviation.

3.2.5 Histone H3-K9/K14 acetylation across the MHC

In order to profile histone H3-K9/K14 acetylation across the MHC, DNA samples enriched in H3-K9/K14 acetylation were obtained by ChIP and then hybridised to the MHC tiling path arrays. HT1080 fibroblasts were treated with formaldehyde and subjected to standard ChIP with antibodies directed against H3-K9/K14 acetylation. ChIP was also performed using non-specific mouse IgG antibodies as a negative control. Acetylation of H3-K9/K14 has been previously found at the 5' regions of transcriptionally active human genes (Liang et al., 2004; Roh et al., 2005). To assess the quality of the ChIP, H3-K9/K14 acetylation was analysed with quantitative real-time PCR at the 5' regions of *HLA-A*, *HLA-C*, *HLA-DRA* and *HLA-DRB1*. In HT1080 fibroblasts, the *HLA-A* and *HLA-C* genes are constitutively expressed while *HLA-DRA* and *HLA-DRB1* genes are transcriptionally silenced (Rohn et al., 1996; Singer and Maguire, 1990). In agreement with previous reports, the PCR data show that *HLA-A* an *HLA-C* genes are significantly more H3-K9/K14 acetylated compared to *HLA-DRA* and *HLA-DRB1* genes confirming that H3-K9/K14 acetylation is associated with active transcription of human genes (Fig 3.6) (Agalioti et al., 2002; Liang et al., 2004).

The DNA immunoprecipitates obtained using anti-acetyl H3-K9/K14 and mouse IgG antibodies were labelled with Cy5 and Cy3 fluorophores and co-hybridised to the MHC tiling path array. A dye swap experiment was carried out to reduce the bias introduced by the different fluorescent labelling of the DNA samples. The ratios of the fluorescent dyes of each locus were $\log_2$ transformed and the combined mean values calculated. An arbitrary cut-off of array enrichment was set at 3 standard deviations of the mean to identify H3-K9/K14 acetylated MHC regions which were subsequently plotted on the UCSC human genome database (Fig 3.7). The results show that 281 regions of the MHC, representing 17% of the array probes, are associated with K9/K14 acetylated H3. These enriched probes are unevenly distributed across the MHC locus. Over 47% of the H3-K9/K14 acetylated regions are contained in the MHC class I, 31% in the MHC class III, 8% in the MHC class II and 14% in the extended MHC class II regions.

To determine if the H3-K9/K14 acetylation is biased towards any functional regions in the human genome, the MHC was divided into transcription initiation sites, gene body regions including introns and exons, and intergenic regions. Approximately 57% of the array probes covering the MHC locus contain entirely intergenic regions, 31% contain

gene body regions while 12% contain transcription initiation sites (Fig 3.8a). This distribution is significantly different from that of the array probes enriched in H3-K9/K14 acetylation. Of such probes 23% contain intergenic regions, 43% contain gene body regions and 34% contain transcription start sites (Fig 3.8b). H3-K9/K14 acetylation is thus biased towards the transcription sites and gene body regions of the genome. The distribution of the 50 array probes most enriched in H3-K9/K14 acetylation is even more shifted towards the transcription initiation sites (Fig 3.8c). This suggests that the 5' ends of genes are the most highly acetylated regions, as previously shown by a genome-wide analysis of H3-K9/K14 acetylation in T cells (Roh et al., 2005).

In agreement with the real-time PCR results, the array data show that the regions overlapping the 5' region of *HLA-A* and *HLA-C* are enriched in H3-K9/K14 acetylated DNA while *HLA-DRA* and *HLA-DRB1* are not H3-K9/K14 acetylated. This indicates that the arrays are suitable for the profiling of histone modifications across the MHC locus. The genes that are H3-K9/K14 acetylated include all the classical MHC class I genes but none of the classical MHC class II genes (Table 3.4). While fibroblasts constitutively express the classical MHC class I genes, they do not express the classical MHC class II genes unless induced with cytokines such as IFN-γ (David-Watine et al., 1990; Rohn et al., 1996; Singer and Maguire, 1990). These experiments further validate the MHC tiling path arrays since they are consistent with previous reports associating transcriptionally active genes with H3-K9/K14 acetylation (Agalioti et al., 2002; Liang et al., 2004; Roh et al., 2005).

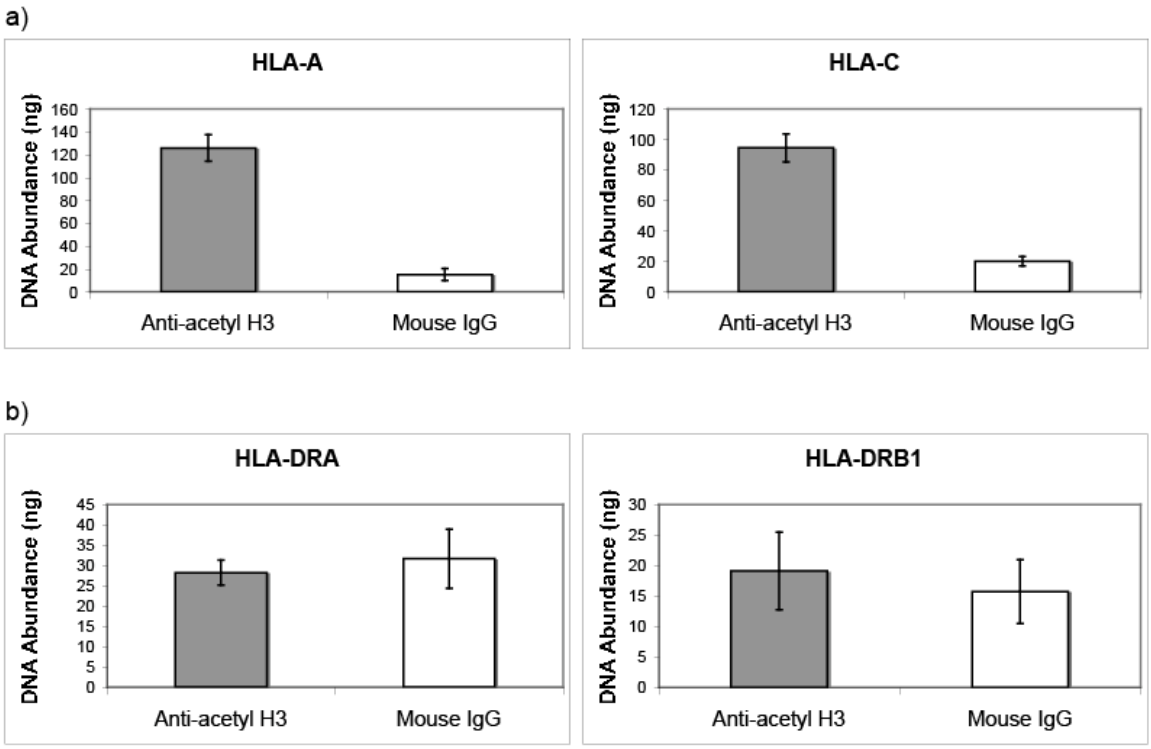

Figure 3.6 Quantitative real-time PCR analysis at the 5' regions of *HLA-A*, *HLA-C*, *HLA-DRA* and *HLA-DRB1* in HT1080 cells.

Real-time PCR was carried out using two samples each of ChIP DNA from HT1080 cells and primer pairs at the 5' regions of *HLA-A*, *HLA-C*, *HLA-DRA* and *HLA-DRB1*. **a)** The data indicate that the 5' regions of *HLA-A* and *HLA-C* are enriched in the ChIP samples obtained by using anti-acetyl H3-K9/K14 compared to the samples obtained by using mouse IgG. **b)** No significant differences are observed between the two ChIP samples relative to the enrichment of the 5' region of *HLA-DRA* and *HLA-DRB1*. Bars represent the mean of two experiments and show the standard deviation.

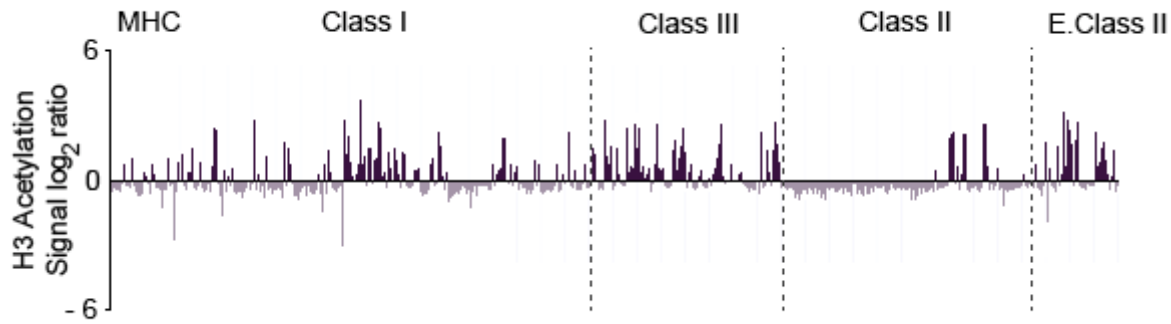

Figure 3.7 H3-K9/K14 acetylation profile of the MHC.

The plot was obtained by co-hybridising to the MHC tiling path array Cy3- and Cy5-labelled ChIP DNA from HT1080 cells using anti acetyl-histone H3-K9/K14 and mouse IgG. The values above 0 represent enrichment in acetyl-H3-K9/K14 ChIP DNA. The values below 0 represent enrichment in mouse IgG ChIP DNA. The experiment was carried out twice using ChIP DNA from two biologically independent experiments. The height of each bar shows the average $\log_2$ enrichment.

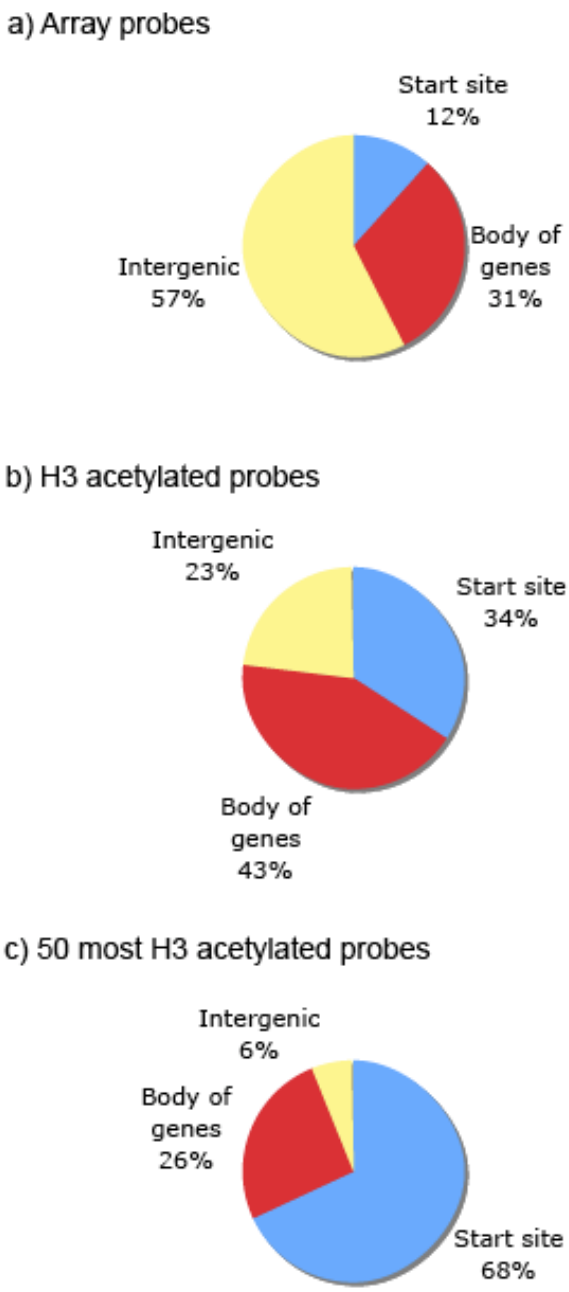

Figure 3.8 Distribution of H3-K9/K14 acetylated probes across the start sites of genes, within genes and intergenic regions of the MHC in HT1080 cells.

Array probes were divided into three groups: overlapping transcription initiation sites, gene body regions including introns and exons and intergenic regions. The pie charts show the calculated percentages of each group in **a)** MHC array probes, **b)** H3-K9/K14 acetylated probes and **c)** the 50 most acetylated H3-K9/K14 probes.

Number	Gene	Chr. co-ordinates	Number	Gene	Chr. co-ordinates
1	*MOG*	29732788-29748126	48	*BAT4*	31737842-31741142
2	*HLA-F*	29799096-29803052	49	*CSNK2B*	31741636-31745822
3	*HLA-G*	29902735-29906878	50	*LY6G5B*	31746707-31748206
4	*HLA-A*	30018310-30021632	51	*LY6G5C*	31794405-31797489
5	*HCG9*	30050871-30054148	52	*BAT5*	31762715-31779067
6	*ZNRD1*	30137015-30140665	53	*DDAH2*	31802797-31806018
7	*PPP1R11*	30142911-30146085	54	*CLIC1*	31806339-31812320
8	*TRIM31*	30178655-30188846	55	*MSH5*	31815753-31838429
9	*TRIM15*	30238962-30248452	56	*LSM2*	31873155-31882722
10	*TRIM26*	30260213-30289132	57	*HSPA1L*	31885375-31890814
11	*TRIM39*	30403019-30419484	58	*HSPA1A*	31891316-31893696
12	*RPP21*	30420916-30422611	59	*HSPA1B*	31903491-31906010
13	*HLA-E*	30565250-30569064	60	*C6ORF48*	31910672-31915518
14	*GNL1*	30621675-30632987	61	*NEU1*	31934932-31938662
15	*PRR3*	30632735-30640159	62	*SLC44A4*	31938949-31954802
16	*ABCF1*	30647149-30667286	63	*ZBTB12*	31975373-31977748
17	*PPP1R10*	30676162-30692999	64	*CFB*	32021752-32027839
18	*MRPS18B*	30693465-30702153	65	*RDBP*	32027844-32034843
19	*C6ORF134*	30702592-30720427	66	*SKIV2L*	32034560-32045511
20	*C6ORF136*	30722780-30728961	67	*STK19*	32047625-32057201
21	*DHX16*	30728886-30748736	68	*C4A*	32090550-32111173
22	*NRM*	30763806-30766748	69	*TNXB*	32116911-32185129
23	*MDC1*	30775563-30793437	70	*CREBL1*	32191023-32203995
24	*TUBB*	30796136-30801172	71	*FKBPL*	32204465-32206045
25	*FLOT1*	30803492-30818432	72	*PPT2*	32229279-32239429
26	*IER3*	30818955-30820306	73	*AGPAT1*	32243968-32253820
27	*DDR1*	30959840-30975910	74	*AGER*	32256724-32260001
28	*GTF2H4*	30983956-30989858	75	*PBX2*	32260488-32265941
29	*VARS2*	30989961-31002212	76	*GPSM3*	32266522-32271278
30	*PSORS1C1*	31190602-31215816	77	*NOTCH4*	32270599-32299822
31	*PSORS1C2*	31213290-31215106	78	*TAP2*	32897588-32914525
32	*CCHCR1*	31218195-31233545	79	*PSMB8*	32916477-32919794
33	*TCF19*	31234282-31239970	80	*TAP1*	32920965-32929726
34	*HCG27*	31273578-31279723	81	*BRD2*	33044415-33057260
35	*HLA-C*	31344509-31347834	82	*RXRB*	33269344-33276410
36	*HLA-B*	31429628-31432968	83	*SLC39A7*	33276581-33280191
37	*MICA*	31479350-31491069	84	*RING1*	33284264-33288476
38	*HCP5*	31538938-31541460	85	*VPS52*	33326028-33347640
39	*MICB*	31573834-31586880	86	*RPS18*	33347830-33352259
40	*BAT1*	31605977-31618204	87	*PFDN6*	33365356-33366689
41	*ATP6V1G2*	31620221-31622606	88	*RGL2*	33367416-33374716
42	*NFKBIL1*	31623351-31634584	89	*TAPBP*	33375452-33389967
43	*LTB*	31656317-31658181	90	*DAXX*	33394379-33398682
44	*LST1*	31661950-31664664	91	*KIFC1*	33467291-33485677
45	*BAT2*	31696429-31713533	92	*PHF1*	33486772-33492191
46	*BAT3*	31714784-31728456	93	*CUTA*	33492297-33494043
47	*C6ORF47*	31734054-31736528	94	*SYNGAP1*	33495825-33529443

Table 3.4 H3-K9/K14 acetylated genes across the MHC in HT1080 cells identified by ChIP assays coupled with MHC tiling path arrays. The H3-K9/K14 acetylated genes were identified setting an array enrichment cut-off of 3 standard deviations of the mean.

3.3 Discussion

The MHC tiling path arrays were validated using three strategies: 1) self-self DNA hybridisations 2) hybridisations of CIITA-enriched DNA 3) hybridisations of DNA enriched in H3-K9/K14 acetylation. Self-self hybridisations indicated that the array probes exhibited a high degree of hybridisation uniformity. Hybridisations of DNA enriched in CIITA and H3-K9/K14 acetylation showed that specific DNA-protein interactions at previously characterised MHC sites could be detected with high reliability. For each set of experiments, findings compared well with quantitative real-time PCR. Taken together, these results demonstrate the utility of the tiling path array in profiling DNA-protein interactions across the MHC locus.

In contrast to most commercial arrays, these MHC array platforms include repetitive sequences, which were blocked in each experiment using Cot1 DNA. Control hybridisations were performed with the MHC tiling path arrays, with and without Cot1 DNA (Tomazou et al., 2008). These experiments showed that when using Cot1 DNA, the intensities of repeat-containing probes are similar to those detected for repeat free probes. This suggests that unwanted signals derived from repetitive elements can be efficiently suppressed and that the unique regions of repeat containing probes are informative and can be used for further analysis (Tomazou et al., 2008). The array probes constituted entirely of repetitive sequences were removed from further analyses since hybridisation to those sequences would be non-specific.

Control self-self hybridisations were carried out to identify any array probes that failed to give the expected ratios of fluorescent dyes and to determine the experimental variation. In these experiments, the $\log_2$ ratio for each probe was expected to be close to zero. The results did not reveal any probe with a ratio deviating significantly from the expected values. The experimental variation, defined as the technical background signal, was determined by calculating the standard deviation of the array probes. This was extremely low in each hybridisation indicating that the results were consistent and reproducible. The results of the self-self experiments were then validated using an independent method. Quantitative real-time PCR was carried out across six randomly selected regions, indicating that there were no significant enrichments of either DNA sample, thus further confirming the array data.

Array hybridisations with ChIP DNA samples obtained using antibodies against CIITA identified 27 CIITA-enriched loci. These sequences include 12 proximal-promoter regulatory modules, SXY, and 2 distal regulatory modules, S'-Y', of the classical and non-classical MHC class II genes. The MHC class II contains a total of 12 proximal-promoter regulatory modules, SXY, and 4 distal regulatory modules, S'-Y', for a total of 16 characterised CIITA target sites (Krawczyk et al., 2004). The S'-Y' module from *HLA-DRA* was not represented on the arrays since it was refractory to PCR amplification. Thus, out of the 15 CIITA target sites spotted onto the arrays, only one S'-Y' module was not detected. This indicates that the MHC tiling path array is suitable for mapping DNA-protein interactions with high sensitivity.

These SXY and S'-Y' modules bind to RFX, CREB and NF-Y to form a landing platform that recruits CIITA (Krawczyk et al., 2004; Louis-Plence et al., 1997; Masternak et al., 2000). These proximal-promoter modules have been shown to be essential for the expression of the classical and non-classical MHC class II molecules (Reith and Mach, 2001). The promoters of the genes coding for the classical and non-classical MHC class I molecules are also characterised by the SXY module (Gobin and van den Elsen, 2000; Krawczyk et al., 2008; van den Elsen et al., 2004). Although CIITA is involved in the basal and induced transcriptional activation of these MHC class I genes, its binding to their regulatory sequences has not been yet shown (van den Elsen et al., 2004). The MHC tiling path arrays coupled with the ChIP assay indicated that the array probes overlapping the SXY modules of several genes of the MHC class I are associated with CIITA. This suggests that CIITA might play a role in the expression of classical and non-classical MHC class I molecules by interacting at conserved binding sites.

Six novel potential CIITA-binding sites were identified. These either overlap or are proximal to genes which have been implicated in a wide range of human diseases. *TRIM14* encodes a member of the tripartite motif (TRIM) family of ubiquitin ligases (Meroni and Diez-Roux, 2005). By promoting the ubiquitination of substrate proteins, members of this family can alter their cellular location or function as well as regulating their degradation via the proteosome (Krawczyk et al., 2008). Krawczyk et al. (2008) suggested that an altered expression of TRIM proteins could potentially contribute to the development of tumours (Krawczyk et al., 2008). *MICA* encodes a transmembrane glycoprotein that functions as a ligand for NK cells (Zou et al., 2005). The expression of

MICA has been associated with a risk of developing several autoimmune and immune-mediated diseases (Gambelunghe et al., 2005). *DPCR1* synthesises a multiple-domain transmembrane protein which has been linked with diffuse panbronchiolitis (Matsuzaka et al., 2002). *HLA-G* is an HLA class I heavy chain paralogue. Its expression is confined to extravillous cytotrophoblast cells and has been proposed to play a role in materno-fetal tolerance (Gobin and van den Elsen, 1999). Previous studies suggest that *HLA-G* transcription is not responsive to CIITA mediated pathways, however, the implication of CIITA in the regulation of *HLA-G* cannot be conclusively excluded at present (Gobin and van den Elsen, 1999). *TUBB* synthesises β-tubulin, a GTPase. Mutations in the *TUBB* gene were shown to be a good predictor of chemotherapy responses in non-small cell lung cancer (Monzo et al., 1999). *TAP1* encodes a membrane-associated protein that is involved in the processing of endogenous peptides that bind to MHC class I molecules (Alvarado-Guerri et al., 2005). This protein is associated with susceptibility to coeliac disease and insulin-dependent diabetes mellitus (Colonna et al., 1992).

A recent microarray study by Krawczyk et al. (2008) did not identify these six potential CIITA target sites. This work was carried out with a microarray carrying only the promoter regions of 27434 human genes (Krawczyk et al., 2008). Thus the novel putative CIITA-binding sites identified in *TUBB* and *MICA* could have not been identified as they are located away from their promoter regions. The remaining four potential CIITA target sites could have been missed since Krawczyk et al. (2008) used a different microarray platform (NimblGen) and analysed the data using a software (SignalMap, NimblGen) which identified positive peaks using a 500 bp sliding window containing five consecutive peaks. It is also possible that these novel potential CIITA-bindings sites are false positives generated during the amplification step that I used to generate sufficient quantities of DNA for microarray analysis. The DNA was amplified using ligation-mediated PCR, an approach which could introduce sequence-dependent and length-dependent biases due to the ligation efficiency and kinetics of the PCR reaction. Furthermore, the amplification of low abundance species can be highly variable. One way to validate the CIITA target sites identified with the microarrays would be to identify them with PCR using ChIP DNA products which have not been subjected to amplification. Due to the amplification step, the quantitative real-time PCR could not conclusively confirm the novel CIITA target sites, however, it was successful

in validating the array results demonstrating that the MHC tiling path array can be used to screen DNA-protein interactions.

To further validate the array platform we used ChIP in combination with the MHC tiling path arrays to profile histone H3-K9/K14 acetylation. This histone modification is a hallmark of transcriptional activation and has been previously reported to be associated with the 5' ends of active human genes (Kurdistani et al., 2004; Liang et al., 2004; Pokholok et al., 2005; Roh et al., 2005). The array results have shown that the 5' ends of the constitutively active classical MHC class I genes are highly acetylated while the 5' ends of the transcriptionally silent classical MHC class II genes do not exhibit high levels of acetylation. This experiment did not require an amplification step since the DNA quantities obtained from ChIP using antibodies against acetylated H3-K9/K14 were suitable to carry out array hybridisations. Quantitative real-time PCR confirmed the array results demonstrating that the MHC tiling path array is a powerful tool for the study of the histone modifications characterising the MHC locus. Major advances in our understanding of the role of histone acetylation in regulating gene expression comes from studies in yeast where the genome has high levels of acetylation, largely localised at the promoter as well as the coding regions of active genes (Bernstein et al., 2002; Kurdistani et al., 2004). Recent studies in human cells, however, have suggested that H3-K9/K14 acetylation occurs predominantly at the beginning of transcriptionally active genes (Liang et al., 2004; Litt et al., 2001; Roh et al., 2005). This is in agreement with the MHC array data, which indicate that acetylation is highest at the 5' ends of constitutively expressed MHC genes and decreases downstream of their start sites.

A recent study has shown that the MHC tiling path array is suitable for the profiling of DNA methylation across the MHC of different cell types (Tomazou et al., 2008). This investigation, in combination with the study of H3-K9/K14 acetylation of the MHC, strongly suggests that this array platform can facilitate comprehensive epigenetic analyses of this region. In addition, the experiments identifying CIITA target sites indicate that these arrays can be also used to map MHC regions interacting with proteins involved in the regulation of MHC genes. Based on the design of these arrays we also expect them to be compatible with additional assays including array comparative genomic hybridisation (aCGH), expression profiling and the biochemical extraction of MARs.

4 Genomic anchors in MRC5 fibroblasts

4.1 Introduction

MARs have been implicated in numerous aspects of gene regulation (Berezney et al., 1995; Singh et al., 1997). They can recruit chromatin remodelling enzymes, histone modifying enzymes, and transcription factors, including SATB1 and SAFB1, and can increase transgene expression by shielding the transgene from the silencing effects of neighbouring chromosomal regions (Cai et al., 2006; Girod et al., 2007; Oesterreich, 2003). However, before this study was performed only a few MARs were identified, limiting our understanding of how they function (Linnemann et al., 2007).

Individual MARs were identified previously by quantitative real-time PCR and *in vitro* binding affinity of genomic sequences to the nuclear matrix. The need to identify MARs on a genome-wide scale has led to the development of algorithms for their prediction based on the sequence patterns associated with experimentally defined MARs. However, the predicted sequences are not always in agreement with experimental data (Linnemann et al., 2007). Such inconsistencies could be due to our current incomplete understanding of MAR sequences and/or to the fact that a significant proportion bind to the nuclear matrix in a cell and stage-specific manner (Heng et al., 2004; Linnemann et al., 2007; Ostermeier et al., 2003).

The development of high-throughput technologies presents the opportunity to construct detailed genome-wide maps of MARs. Ioudinkova et al. (2005), working on the 40 kb domain at the α-globin gene domain, were the first to show that MARs could be successfully identified by hybridising them onto oligonucleotide arrays of a specific genomic region (Ioudinkova et al., 2005). Taking a different approach, Linnemann et al. (2008) identified MARs on human chromosome 16 by competitively hybridising MAR and halo DNA onto genomic microarrays (Linnemann et al., 2008). The positions of MAR and halo DNA sequences, which were previously differentially labelled with fluorochromes, were then determined by calculating the ratio of fluorochrome intensities at each microarray probe.

Since the primary sequence and linear organisation of the MHC have been extensively

characterised, it provides an ideal model region to investigate the structural and functional role of MARs. In this chapter, MARs were mapped across the MHC of human fibroblasts using the biochemical isolation of MAR and halo DNA fractions followed by their co-hybridisation to the tiling path array covering the MHC locus.

MARs were found to be non-randomly distributed across the MHC, since most were located in the class I and II regions. Furthermore, they mainly had intergenic positions, closely flanking several class I and II genes as well as the MHC class boundaries. These data suggest that MARs define chromatin domains and contribute to the class-specific regulation of the locus. Sequence analysis of MARs did not reveal a unique consensus but established significant correlations between MARs and particular genomic features, reinforcing the view that the DNA binding ability to the nuclear matrix might depend on DNA sequence context rather than on a unique primary sequence. Finally, validation of the microarray data by real-time PCR indicates that the array platform can reliably identify interactions of MHC sequences with the nuclear matrix and can thus be used to characterise the dynamics of MARs in cells exhibiting different MHC expression profiles. The findings in this Chapter and Chapter 5 are included in a paper in press (Ottaviani et al., 2008a).

4.2 Results

4.2.1 Biochemical isolation of matrix attachment regions from MRC5 fibroblasts

MARs were isolated by treating nuclei from MRC5 fibroblasts with 2 M NaCl followed by nuclease digestion. The hypertonic treatment removed the soluble proteins, including histones, causing the DNA to loop out of the nuclei (Cai and Kohwi-Shigematsu, 1999; Kramer and Krawetz, 1996). This resulted in the formation of histone-depleted halo structures consisting of extruded 'loop' DNA. Within the residual nucleus, the DNA is anchored to the nuclear matrix through MARs (Gerdes et al., 1994; Kramer and Krawetz, 1996). After digesting these structures with restriction enzymes, sedimentation was used to separate the MARs from the halo DNA (Ostermeier et al., 2003).

To allow the full extraction of soluble proteins from nuclei and obtain reproducible results, the time of incubation with 2 M NaCl had to be optimised. The optimal time, defined as the minimum incubation time required to reach the maximum halo size, was previously shown to vary according to the cell type (Martins and Krawetz, 2007; Ostermeier et al., 2003).

Nuclear preparations were made from MRC5 cells, placed onto microscope slides and incubated in 2 M NaCl for varying times ranging from 0 to 15 min. The nuclei were then stained with DAPI and visualised under a Zeiss Axiophot microscope equipped for epifluorescence. The average size of the halo structures was determined by measuring the distance between the perimeter of the halo and the edge of the residual nuclear border of 30 cells in 3 biologically independent experiments. The optimal incubation time for MRC5 nuclei was found to be 12 min as this produced the largest circular halo structures (Fig 4.1, Fig 4.2). The nuclei not incubated with 2 M NaCl, which were used as negative controls, did not form halo structures. Incubation times shorter than the optimal time restricted the extraction of soluble nuclear proteins and led to the formation of small circular nuclear halo structures. Incubation times longer than optimal generated halo structures of highly variable shapes and sizes.

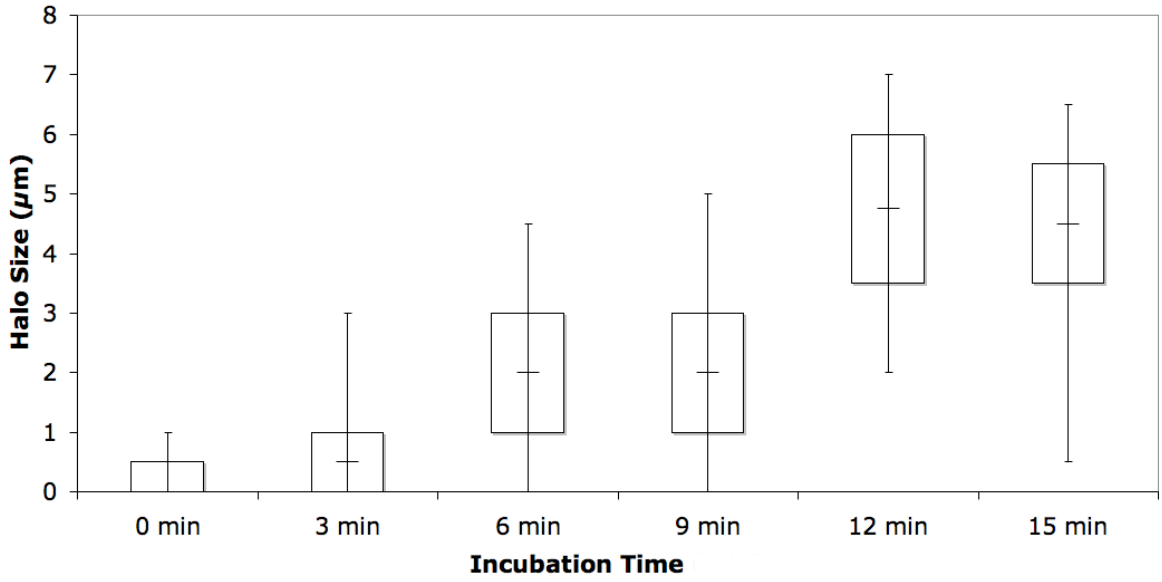

Figure 4.1 Relationship between halo size and incubation time of MRC5 nuclei with 2 M NaCl.

The size of halo structures at each time point was determined by averaging the distance of the perimeter of the halo structure from the edge of the residual nuclear border in 30 nuclei. An increase in halo size between 0 and 12 min of incubation time is followed by a slight decrease at 15 min. The optimal incubation time, defined as the minimum incubation time producing the largest circular halo structures, was 12 min. Boxplots show the interquartile range of the data, with the horizontal black lines running through each boxplot representing the median values of three independent biological experiments. The whiskers extend as far as 1.5 times the interquartile range.

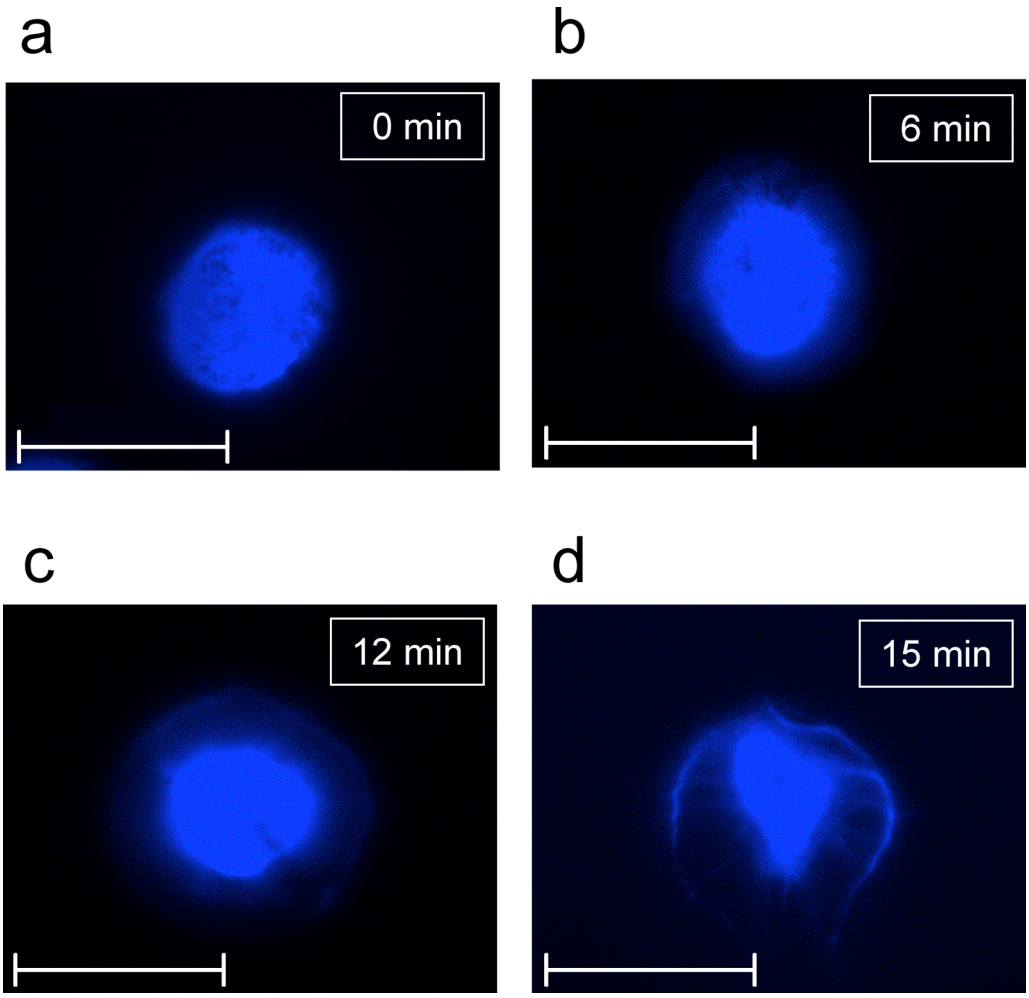

Figure 4.2 Visualisation of nuclear halo structures in nuclei of MRC5 cells.

DAPI-stained nuclei after different incubation times in 2 M NaCl (time shown for each panel). **a)** Halo structures were not seen in untreated nuclei. **b)** Incubation of nuclei with 2 M NaCl for times shorter than optimal generated circular halo structures with a width ranging from 0 to 5 μm. **c)** The optimal incubation time formed circular halo structures with a width ranging from 2 to 7 μm. **d)** Incubations longer than the optimal time produced halo structures of highly variable shapes, with widths ranging from 0.5 to 6.5 μm. The scale bar is 20 μm.

Following optimal extraction of nuclei with 2 M NaCl, EcoRI and HindIII restriction enzymes were added to the extracted nuclei to digest the DNA into short fragments of approximately 1.5-2.5 kb. MARs were protected from digestion by their strong interaction with nuclear matrix proteins and were selectively retained in nuclear matrices, enabling their separation from the dissociated halo DNA. After isolating these two fractions by high-speed centrifugation, their DNA was purified and quantified. The biochemical isolation of MARs was performed five times starting from different MRC5 populations to obtain five independent biological replicates. In all samples, more than 70% of the DNA was found to be in the halo fraction (Table 4.1).

Sample	MAR Fraction (%)	Halo Fraction (%)
1	27.9	72.1
2	19.6	80.4
3	22.4	77.6
4	26.9	73.1
5	27.1	72.9

Table 4.1 Percentage of DNA in the MAR and halo fractions of MRC5 samples.

Over 70% of the DNA was consistently found in the halo fraction, assuming 100% when the MAR and halo fractions were combined.

To confirm that the isolation procedure was performed accurately and consistently, quantitative real-time PCR was performed using the purified MAR and halo fractions to interrogate previously characterised MARs and non-MARs sequences. Each amplicon was quantified using standard curves produced for each primer pair from real-time PCR reactions with known amounts of genomic DNA. To compare the quantity of each amplicon in the MAR and the halo fractions, PCR reactions were carried out with equal

amounts of template and equal amounts of master-mix. For each amplicon, PCR reactions were performed at the same time using the same cycling conditions.

The amplicons PRM2 and F, which map to intragenic regions of the protamine 2 and the δ-globin gene respectively, were used as non-MAR controls since they are sequences consistently enriched in the halo fraction of several somatic cells (Kramer and Krawetz, 1996; Kramer and Krawetz, 1997; Ostermeier et al., 2003). The amplicons HS3 and rep30, which are located within the β-globin LCR and upstream of the γ-globin gene respectively, were used as controls for MARs since they are well characterised in human fibroblasts (Ostermeier et al., 2003). The results show significant enrichment of PRM2 and F in the halo fractions and of HS3 and rep30 in the MAR fractions (Fig 4.3). The proportion of each amplicon in the MAR and halo fractions was calculated by assuming 100% when the average of the MAR and halo fractions were combined. The MAR fractions contained 75% and 72% of the HS3 and rep30 amplicons, respectively, while the halo fractions contained 81% and 83% of the PRM2 and F amplicons, respectively. These findings indicate that the isolation of MAR and halo fractions was consistent, and therefore, these samples could be used for further analysis.

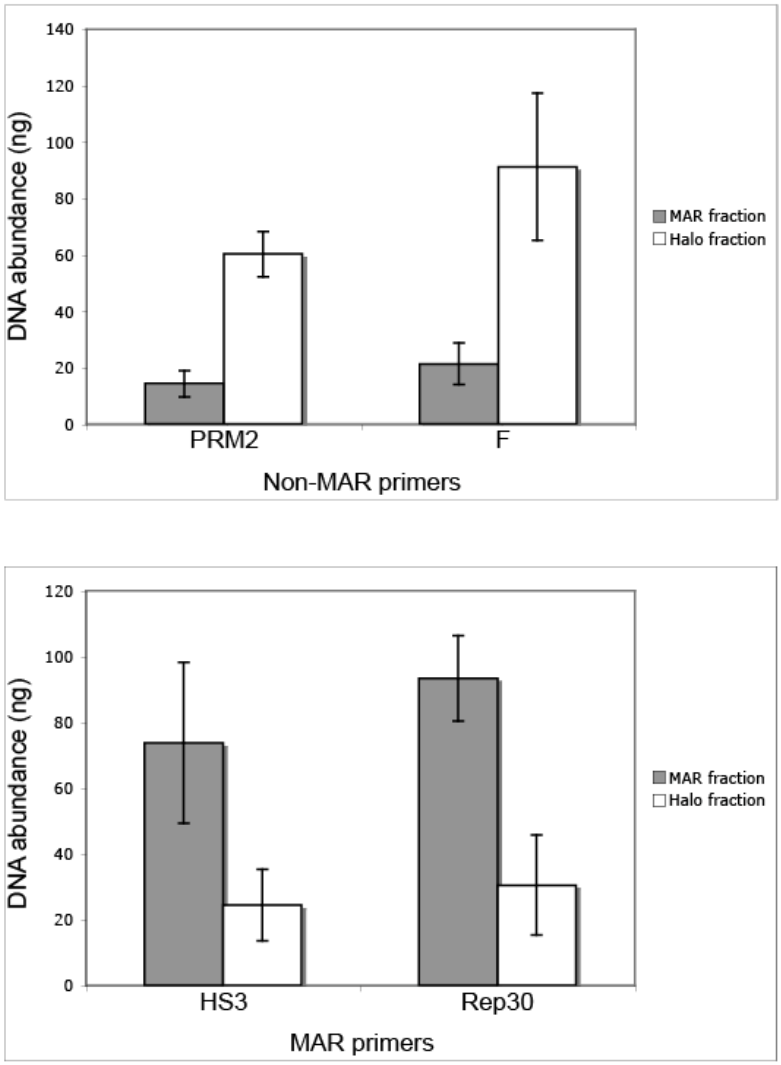

Figure 4.3 Enrichment of MAR and non-MAR amplicons in MAR and halo fractions of MRC5 cells.

Quantitative real-time PCR was performed using matrix and halo DNA samples from five biologically independent experiments and primer pairs spanning known MAR (HS3 and rep30) and non-MAR (PRM2 and F) sequences. Bars represent the mean of five experiments, and show the standard deviation.

4.2.2 Hybridisation of matrix attachment regions and halo DNA onto MHC tiling path arrays

A procedure combining the traditional biochemical isolation of matrix attached DNA with microarray technology was then used to identify MARs across the MHC of MRC5 cells. Equal amounts of MAR and halo DNA from each sample were differentially labelled by random priming with Cy-5 and Cy-3 fluorophores and competitively hybridised onto the genomic MHC tiling path array. The labelled DNA fragments identified the sequences corresponding to MAR and halo DNA in the MHC by binding to matching sequences on the array.

Array hybridisations were carried out with MAR and halo DNA from samples 1 and 2. Both experiments included a dye swap to correct the bias introduced by the different properties of the fluorescent dyes. The relative intensity of the fluorescent dyes across each locus was calculated and $\log_2$ transformed. Each locus was represented four times in the MHC platform. Since two experiments were performed, a total of eight data points were obtained for each locus. Subsequently, the combined mean values for each locus were calculated and plotted on the UCSC human genome database (http://genome.ucsc.edu, May 2004 Assembly) (Fig 4.4).

The graphical representation of the results indicates that MAR and halo DNA sequences are not evenly distributed across the MHC of MRC5 cells. The class I region contains approximately equal amounts of MAR and halo DNA, the class II region contains mainly MAR DNA, and the class III and the extended class II regions consist mainly of halo DNA. The overwhelming majority of MAR DNA is contained across the classical class I and II regions. Visual examination of the results suggests that the halo DNA has a high GC content and is enriched in gene-dense and conserved regions while MARs correlate positively with high AT content and gene-poor regions.

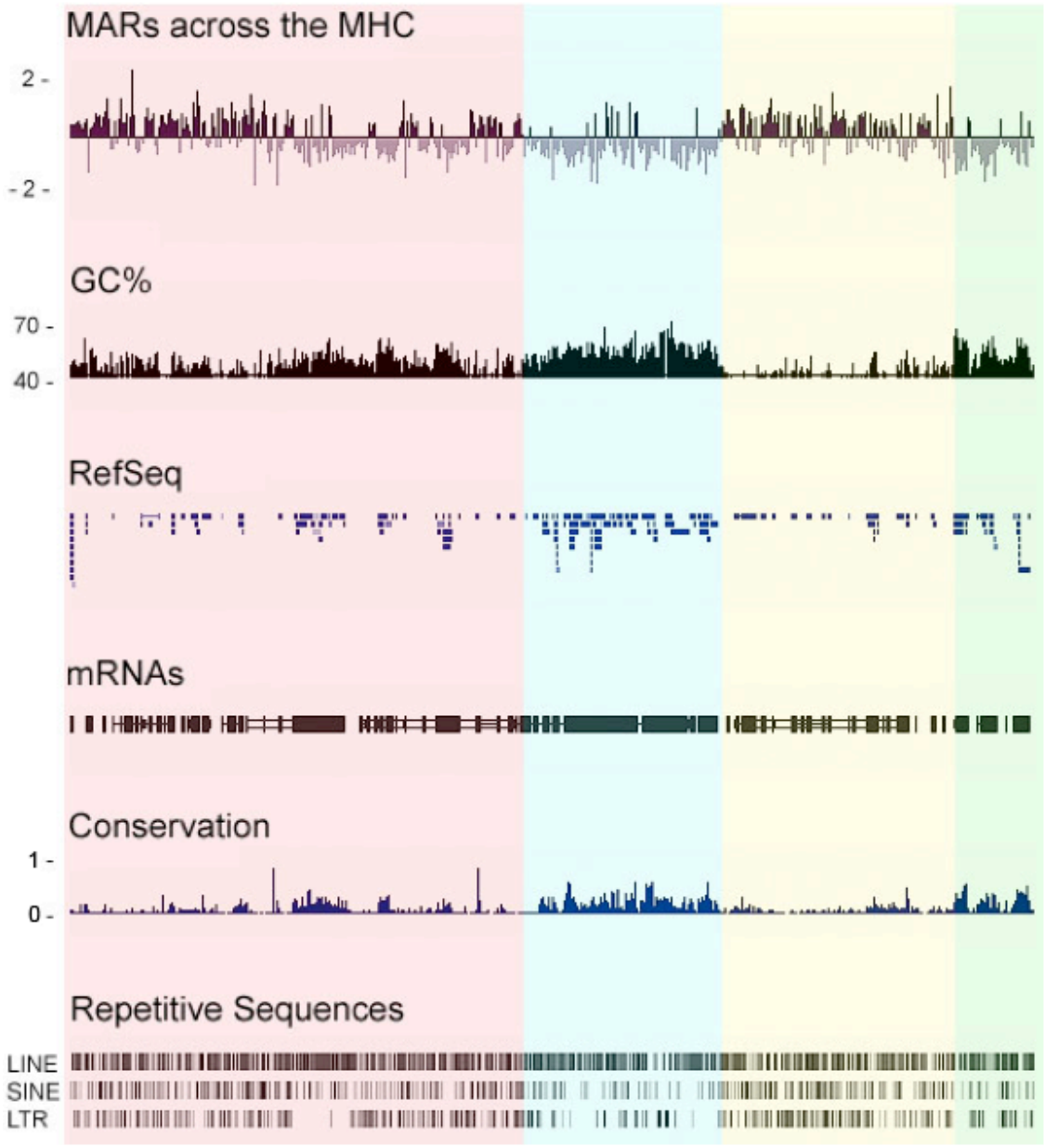

Figure 4.4 Competitive hybridisation of differentially labelled MAR and halo DNA fractions on MHC tiling path arrays.

MAR and halo DNA fractions were differentially labelled with Cy3 and Cy5 fluorophores and competitively hybridised onto the MHC microarray platform. The log2 ratios of MAR/halo DNA were calculated and plotted on the UCSC genome database (http://genome.ucsc.edu/). Positive values represent MARs while negative values represent halo DNA (mean ratios of two independent experiments). The GC content, positions of RefSeq genes and human mRNAs from GeneBank, regions of conservation between 17 vertebrate species and the repetitive sequences LINE, SINE and LTR across the MHC are shown below the array data. The different regions of the MHC are highlighted in different colours: classical class I – red; classical class III – blue; classical class II – yellow; and extended class II – green.

4.2.3 Quantitative real-time PCR analysis of matrix attachment regions and halo DNA

In order to validate the array results, twenty regions across the MHC were subjected to independent quantitative real-time PCR using MAR and halo DNA fractions from samples 3, 4 and 5. Eight regions were enriched in halo DNA, according to the microarray data, while the remaining twelve were MAR-enriched. These test regions were selected within or near genes critical for the functioning of the immune system, genes not directly involved in immunity, as well as boundaries separating the MHC classes.

The eight regions enriched in the halo DNA according to the microarray data were situated within *HLA-G*, *HLA-DRB5*, *HLA-C*, *TAP1*, *HLA-E*, *LY6G5C*, *HLA-DRA* and upstream of *TRIM15*. Real-time PCR confirmed high enrichment in the halo fraction for the regions overlapping *HLA-G*, *HLA-DRB5*, *HLA-C*, *TAP1*, *HLA-E*, *LY6G5C*, and upstream of *TRIM39* and no significant enrichment in the halo fraction for the region overlapping *HLA-DRA* (Fig 4.5). These regions had array ratios equal to or lower than -0.003 ($\log_2$).

Of the twelve regions enriched in the MAR fraction according to the microarray data, eight were successfully validated with quantitative real-time PCR (Fig 4.6). These regions, which were located upstream of *NOTCH4*, *HLA-C*, *ZBTB2* and *HLA-A*, downstream of *HLA-DOB*, *PRRT1* and *HLA-DPB1,* and overlapping *TRIM15*, had array ratios equal to or higher than 0.52 ($\log_2$). The remaining four regions, upstream of *C6orf10*, *HLA-B*, *MICA* and *HLA-F*, were not validated since *HLA-B*, *C6orf10* and *MICA* were highly enriched in the halo fraction and *HLA-F* showed low enrichment in the halo fraction according to real-time PCR (Fig 4.7). These regions had array ratios equal to or lower than 0.32 ($\log_2$).

Further analysis of these results revealed a relationship between high array ratios indicating enrichment in the MAR fraction and successful real-time PCR validation (Table 4.2). In order to exclude false positive MARs, a threshold was empirically determined using a series of real-time PCR as described in Chapter 2. MHC regions having array enrichments equal to or higher than 0.4 ($\log_2$) were defined as MARs since they were consistently confirmed with real-time PCR. MHC regions having array

enrichments lower than 0.4 (log_2) were defined as non-MARs as real-time PCR showed that they were either not significantly MAR-enriched or halo-enriched.

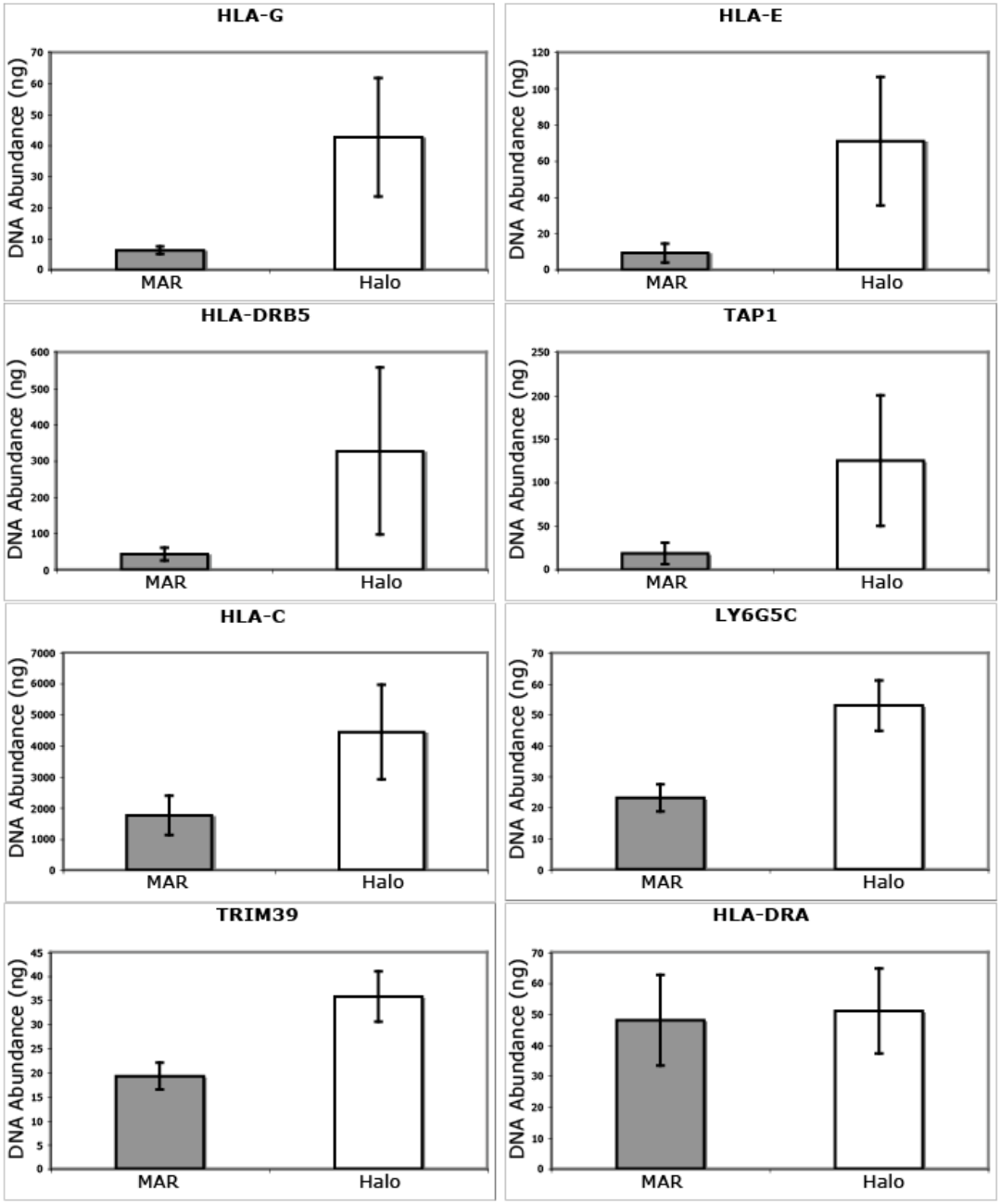

Figure 4.5 Quantitative real-time PCR analysis of halo-enriched loci in MRC5 cells.

Real-time PCR was carried out with primer pairs spanning eight loci, which were enriched in the halo fraction according to the results of array experiments. The loci *HLA-G*, *HLA-DRB5*, *HLA-C*, *TAP1*, *HLA-E*, *LY6G5C* and the upstream region of *TRIM39* were confirmed to be highly enriched in the halo fraction while significant enrichment in the halo fraction was not confirmed for the amplicon overlapping *HLA-DRA*. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

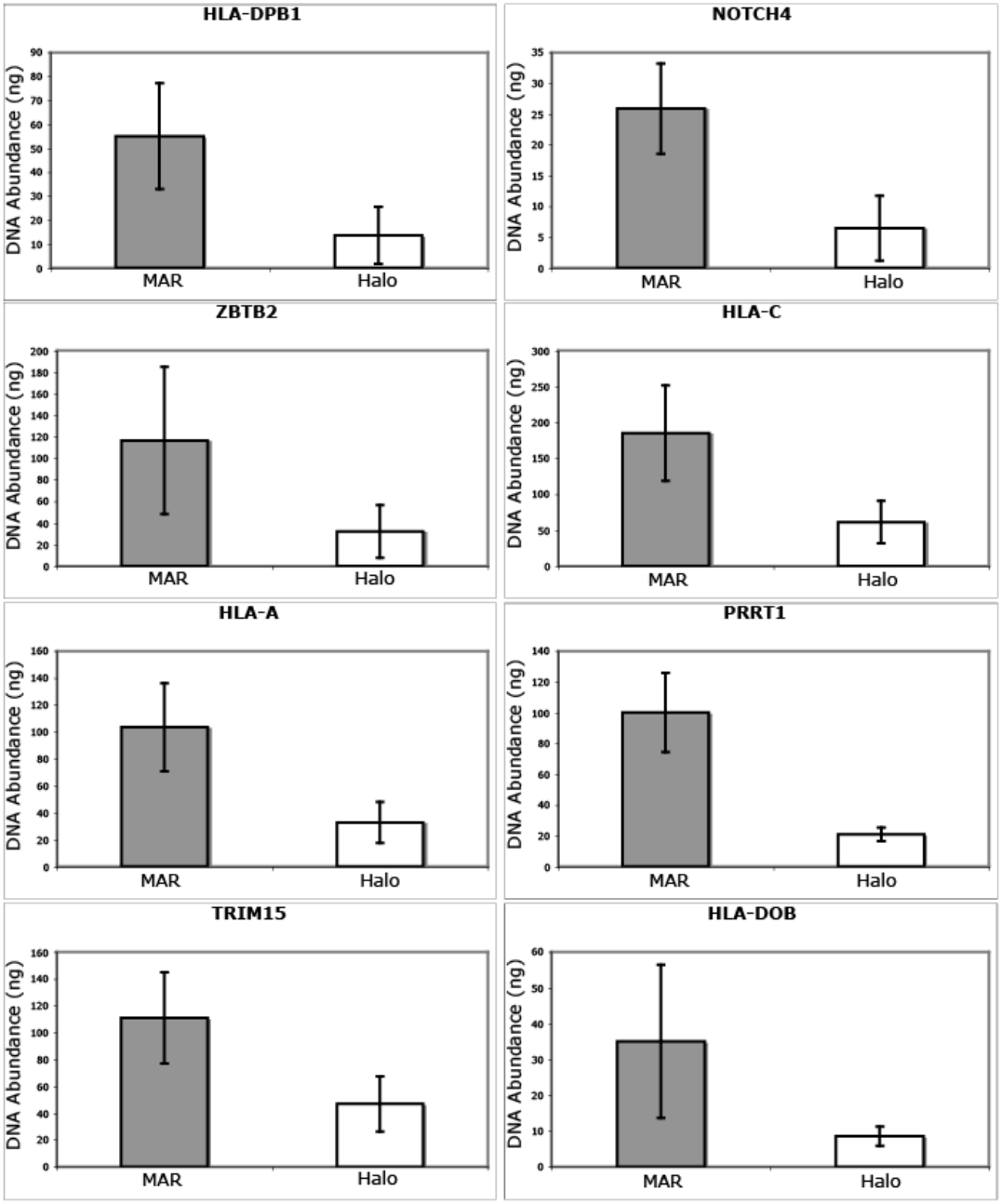

Figure 4.6 Quantitative real-time PCR analysis of loci showing high MAR enrichment in MRC5 cells.

Real-time PCR was carried out with primers pair spanning eight loci, which were found to be highly enriched in the MAR fraction in the array experiments. The regions upstream of *NOTCH4, HLA-C, ZBTB12, HLA-A*, downstream of *HLA-DOB, PRRT1, HLA-DPB1* and overlapping *TRIM15* were confirmed to be highly enriched in the MAR fraction. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

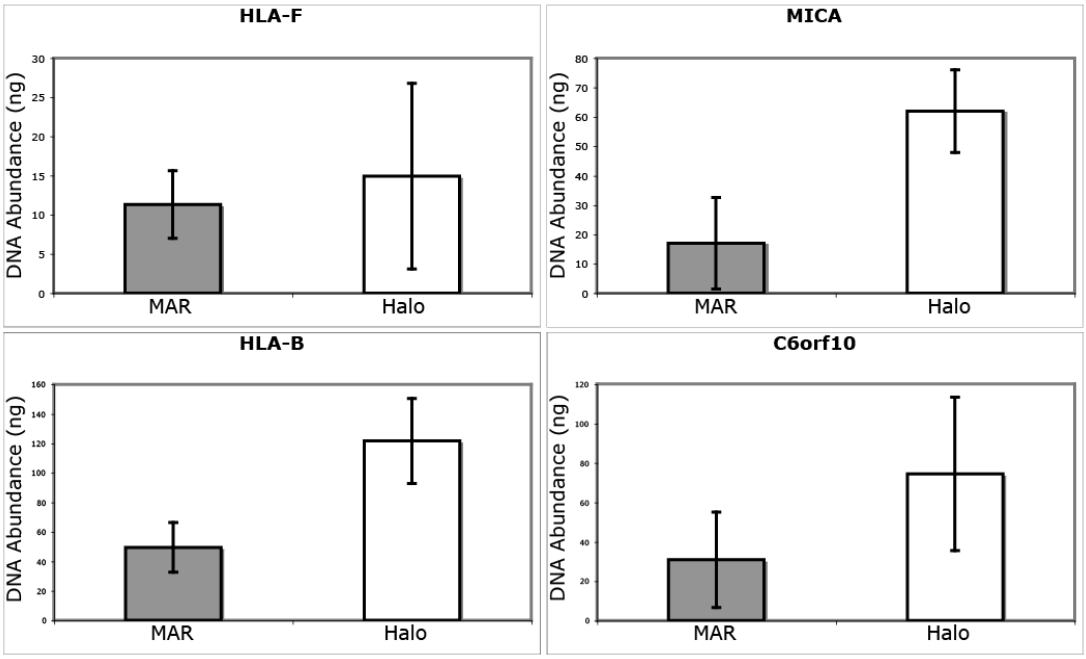

Figure 4.7 Quantitative real-time PCR analysis of loci showing low MAR enrichment in MRC5 cells.

Real-time PCR was carried out with primer pairs spanning within regions of low MAR enrichment according to the array data. The regions upstream of *HLA-B, MICA* and *C6orf10* were confirmed to be highly enriched in the halo fraction while significant enrichment in either the halo or MAR fraction was not observed for the region upstream of *HLA-F*. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

Number	MHC region	Array enrichment value (log2)	Expected enrichment based on array data	Observed enrichment by real-time PCR
1	HLA-A	1.606	MAR	MAR
2	TRIM15	1.104	MAR	MAR
3	HLA-DPB1	1.008	MAR	MAR
4	PRRT1	0.702	MAR	MAR
5	NOTCH4	0.625	MAR	MAR
6	ZBTB2	0.575	MAR	MAR
7	HLA-C	0.522	MAR	MAR
8	HLA-DOB	0.501	MAR	MAR
9	MICA	0.326	MAR	Halo
10	HLA-B	0.29	MAR	Halo
11	C6orf10	0.118	MAR	Halo
12	HLA-F	0.105	MAR	No enrichment
13	HLA-DRB5	-0.003	Halo	Halo
14	TRIM39	-0.1	Halo	Halo
15	TAP1	-0.102	Halo	Halo
16	HLA-E	-0.189	Halo	Halo
17	HLA-DRA	-0.201	Halo	No enrichment
18	HLA-C	-0.202	Halo	Halo
19	HLA-G	-0.27	Halo	Halo
20	LY6G5C	-0.301	Halo	Halo

Table 4.2 Loci analysed by real-time PCR.

(1-8) Regions enriched in the MAR fraction based on the array data and successfully validated with real-time PCR have array ratios ranging from 0.501 to 1.606 (log_2). **(9-12)** Regions enriched in the MAR fraction based on the array data but not able to be validated with real-time PCR have array ratios ranging from 0.326 to 0.105 (log_2). **(12-20)** Regions enriched in the halo fraction according to the array data and successfully validated with real-time PCR have array ratios ranging from -0.003 to -0.301 (log_2). The HLA-DRA and HLA-F regions were not found to be enriched in either the MAR or halo fractions by real-time PCR.

4.2.4 Position of matrix attachment regions across the MHC of MRC5 fibroblasts

To gain a better insight into the distribution of MARs relative to structural and functional characteristics of the MHC locus, a detailed map was constructed for MRC5 cells using the array threshold of 0.4 (log_2) (Fig 4.8).

The graphical representation of the results confirms several trends observed earlier by plotting MAR and halo DNA. MARs are not evenly distributed across the MHC and are largely positioned within poorly conserved regions. The largest clusters of MARs are found within the class I and II regions, while the longest DNA regions lacking MARs are located in the class III and extended class II of the MHC suggesting that MAR position and gene density across the MHC are inversely correlated, as found in AoAF fibroblasts and the HeLa cell line (Linnemann, A. K., personal communication).

Using the criteria described above, MARs were identified on average every 33 kb, consistent with a study on the 16q22 locus in HeLa cells (Shaposhnikov et al., 2007). A total of 117 MARs were identified across the MHC and subdivided as follows: 64 in the class I, 5 in the class III, 43 in the class II and 5 in the extended class II regions of the MHC. The class I, III, II and extended class II regions have, on average, one MAR per 29, 144, 20 and 72 kb, respectively.

Of the 117 MARs detected, 97 are intergenic whereas 19 are intragenic. Using a chi-squared test, it was found that the bias of MARs to intergenic regions was statistically significant (p-value <0.0002; for all statistics a threshold p-value <0.01 was considered significant). Of these, 5 are within introns, 10 overlap introns and exons and 4 map to regions overlapping introns, exons and intergenic DNA (Table 4.3). Due to technical limitations including array resolution and the size of the DNA fragments generated by the biochemical isolation procedure, it was not possible to map these MARs more precisely.

In this study, MARs are identified at the boundaries of the MHC class III and class II regions and the MHC class II and extended class II regions, and 30 kb away from the boundary of the MHC class I and class III regions (Fig 4.9). MARs have also been identified in regions flanking several genes that have critical roles in the adaptive

immune response such as *HLA-A*, *HLA-C*, *HLA-E*, *HLA-G*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1* and *HLA-DQB1* (Fig 4.10) (van den Elsen et al., 1998). These findings are consistent with previous studies showing that MARs can function as boundary elements (McKnight et al., 1992; Namciu et al., 1998).

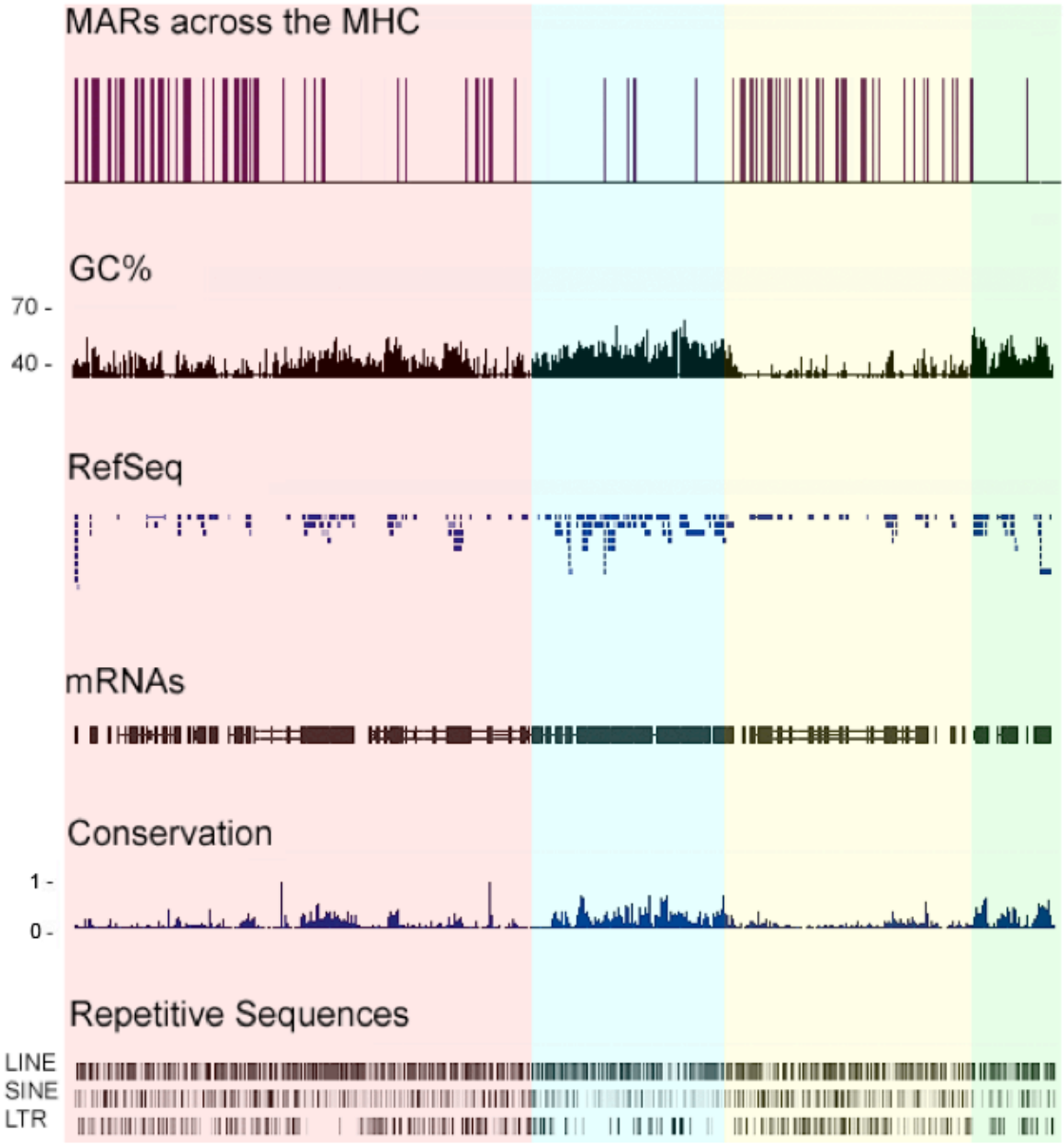

Figure 4.8 MAR profile of MRC5 fibroblasts across the MHC.

The mean value for two array hybridisations was calculated for each fragment represented in the MHC tiling path array. Regions enriched in the MAR fraction with mean ratios equal to or higher than 0.4 ($\log_2$) were plotted on the UCSC genome browser (http://genome.ucsc.edu/) and represented on the top row by black vertical lines. Below the array data are shown the GC content, the positions of RefSeq genes and human mRNAs from GeneBank, regions of conservation between 17 vertebrate species and the repetitive sequences LINE, SINE and LTR. The different regions of the MHC are highlighted in different colours: classical class I – red; classical class III – blue; classical class II – yellow; and extended class II – green.

MAR position (bp)	Gene	Organisation
29802604-29804427	*HLA-F*	1 Exon/1 Intron/Intergenic region
30146157-30148833	*RNF39*	1-2 Exons/1-2 Introns
30231506-30234359	*TRIM10*	4 Exons/3 Introns
30238940-30240972	*TRIM15*	1 Exon/1 Intron
30698815-30701782	*MRPS18B*	2 Exons/2 Introns
30731648-30734384	*DHX16*	4 Exons/5 Introns
31865033-31867206	*VARS*	1 Intron
31955316-31957399	*EHMT2*	3 Exons/3 Introns/Intergenic region
32387922-32390488	*C6ORF10*	1 Intron
32390209-32393140	*C6ORF10*	1 Intron
32395956 32398655	*C6ORF10*	1 Exon/2 Introns
32425489-32427795	*C6ORF10*	1 Exon/1 Intron
32434975-32437634	*C6ORF10*	1 Intron
32471036-32473567	*BTNL2*	1 Exon/2 Introns
32904298-32906120	*TAP2*	3-4 Exons/3 Introns
32931156-32933975	*PMSB9*	4 Exons/3 Introns
33027231-33028946	*HLA-DMA*	1 Exon/1 Intron/Intergenic Region
33283836-33286387	*RING1*	4 Exons/4 Introns/Intergenic Region
33501985-33503886	*SYNGAP1*	1 Intron

Table 4.3 Base pair (bp) position, gene name and DNA organisation of intragenic MARs across the MHC in MRC5 cells.

Out of a total of 19 MARs, 5 are situated within individual introns, 10 are located within regions containing introns and exons and 4 MARs map to regions containing introns, exons and intergenic DNA.

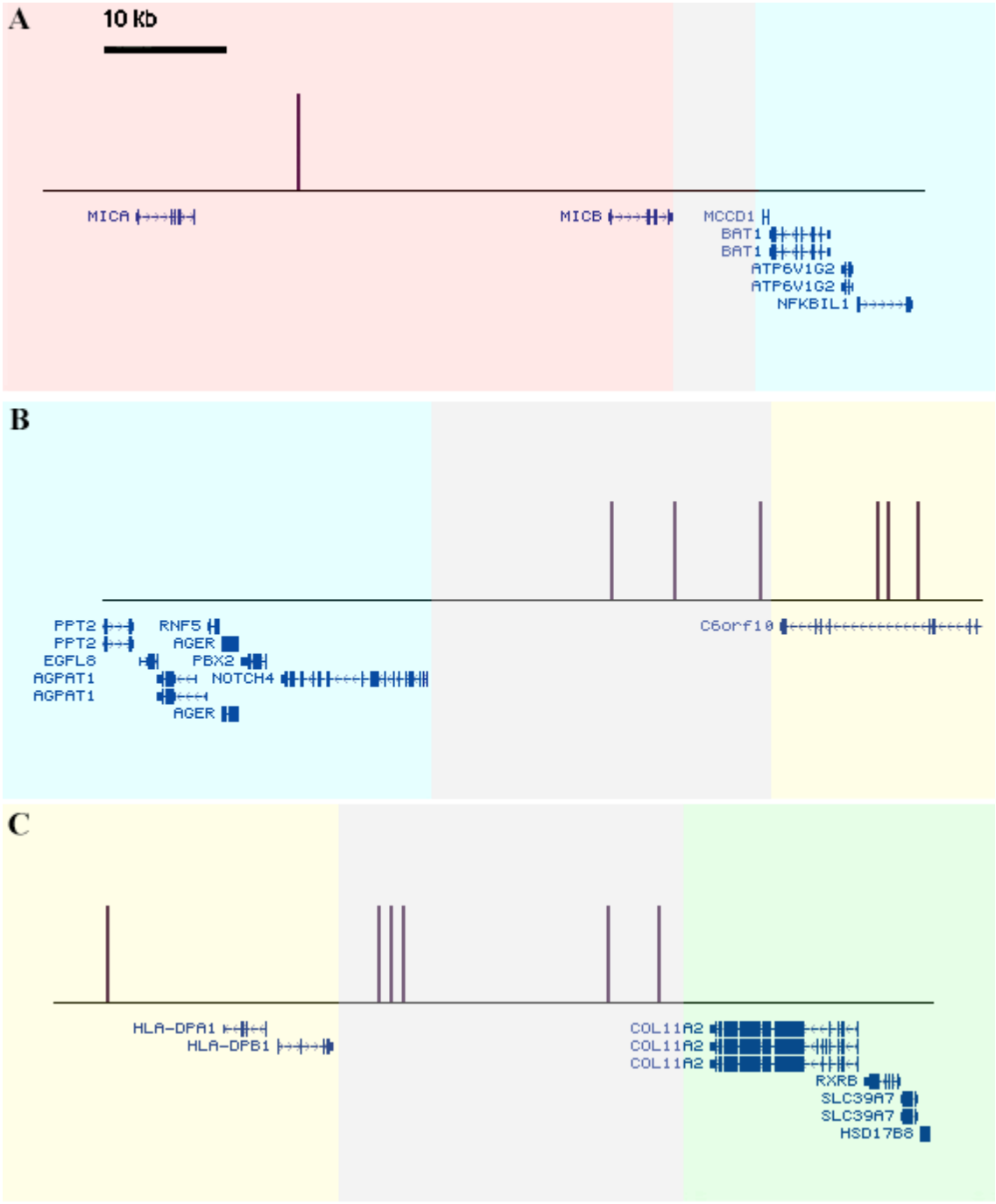

Figure 4.9 Distribution of MARs at MHC class boundaries.

The boundaries separating the MHC into different sub-regions are highlighted in grey. **(A)** The nearest MAR to the boundary separating class I from class III is approximately 30 kb away from it. **(B-C)** The boundaries separating class III from class II and class II from extended class II contain 3 and 5 MARs, respectively.

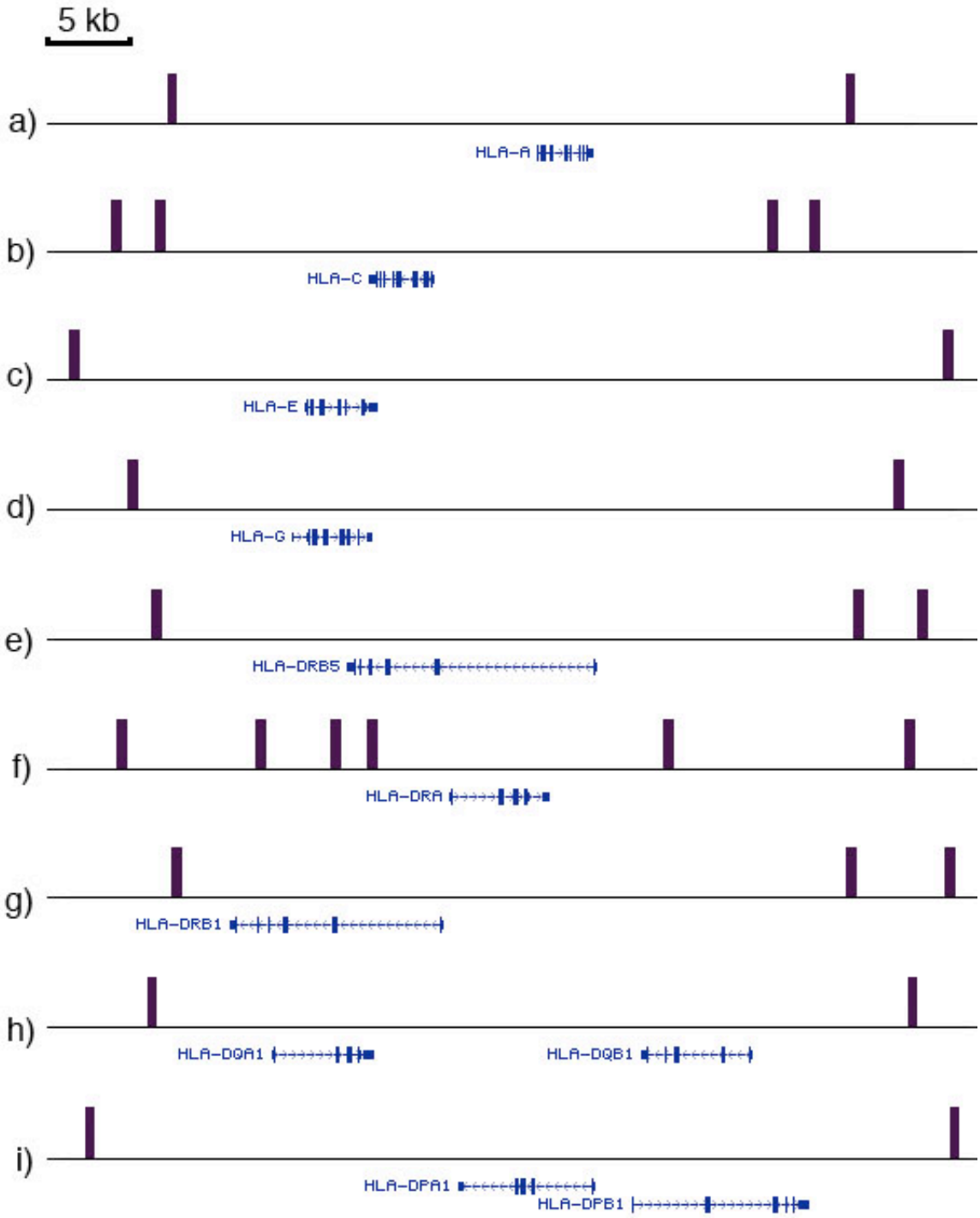

Figure 4.10 Schematic representation of MARs flanking class I and II genes in MRC5 cells.

MARs are represented by vertical black bars. **(a-d)** Class I genes *HLA-A, HLA-C, HLA-E* and *HLA-G*. **(e-i)** Class II genes *HLA-DRB5, HLA-DRA, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1* and *HLA-DPB1*.

4.2.5 Sequence analysis of matrix attachment regions

Several genomic features including conservation, AT content, gene and repetitive sequences density of MAR and non-MARs within the MHC were analysed using public databases and the sequence analysis tools provided by the UCSC Genome Project (http://genome.ucsc.edu/).

The AT content of the MARs ranged from 39 to 67% with an average value of 58% while that of the non-MARs ranged from 32 to 72% with an average value of 53% confirming the previous reports that MARs are generally AT-rich DNA sequences (Fig 4.11a) (Purbowasito et al., 2004; Shaposhnikov et al., 2007). Using a χ^2 test, we found that the difference in AT content between MAR and non-MAR sequences was statistically significant (P-value of 0). Comparative analysis of AT-rich tracts in MARs failed to reveal a precise organisation of AT motifs.

The proportion of intragenic sequences is under-represented by almost twofold among MARs relative to non-MARs (P-value < 0.0002) (Fig 4.11b). This is consistent with the graphical evaluation of the results showing that most MARs identified within the MHC are intergenic and are positioned in relatively gene-poor regions, while non-MARs extend intragenically and contain most gene-rich regions.

According to multiple whole-genome alignments of 17 vertebrates including mammalian, amphibian, bird, and fish species, non-MARs are more conserved than MARs (P-value < 0.008). These findings agree with other studies showing that the primary sequence of MARs is not conserved across species despite the fact that their functions are evolutionary conserved (Fig 4.11c) (Glazko et al., 2003).

The average content of repetitive sequences of MARs is 63%, while that of non-MARs is 45% (P-value of 0) (Fig 4.12a). A detailed analysis of different classes of repetitive sequences revealed that MAR and non-MAR DNA sequences do not differ significantly in relation to LTR content (P-value of 1) (Fig 4.12b), that non-MARs have a larger proportion of SINEs compared to MARs (P-value < 0.006) (Fig 4.12c), and that LINEs are over-represented by almost three-fold among MARs relative to their overall abundance in non-MARs (P-value < 0.002) (Fig 4.12d).

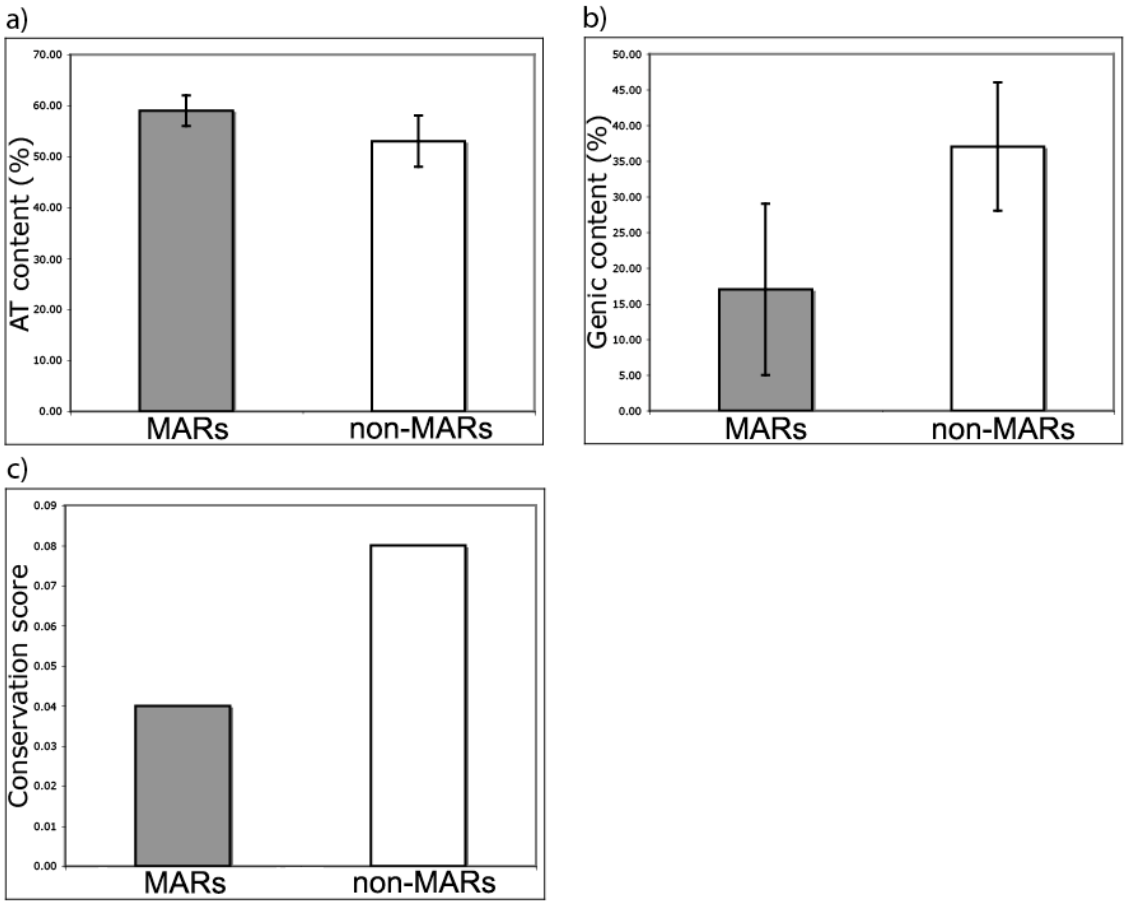

Figure 4.11 Sequence analysis of MARs and non-MARs in relation to different genomic features.

a) The AT content (P-value of 0) **b)** the proportion of coding region (P-value < 0.0002) **c)** and the degree of conservation were calculated for MARs and non-MARs sequences (P-value < 0.008). The conservation score was used as a measure of evolutionary conservation and was calculated using the conservation track of the UCSC Genome Project (http://genome.ucsc.edu/), which describes the multiple alignment of 17 vertebrate species. The bars represent the mean values of two biologically independent experiments and show the standard deviation.

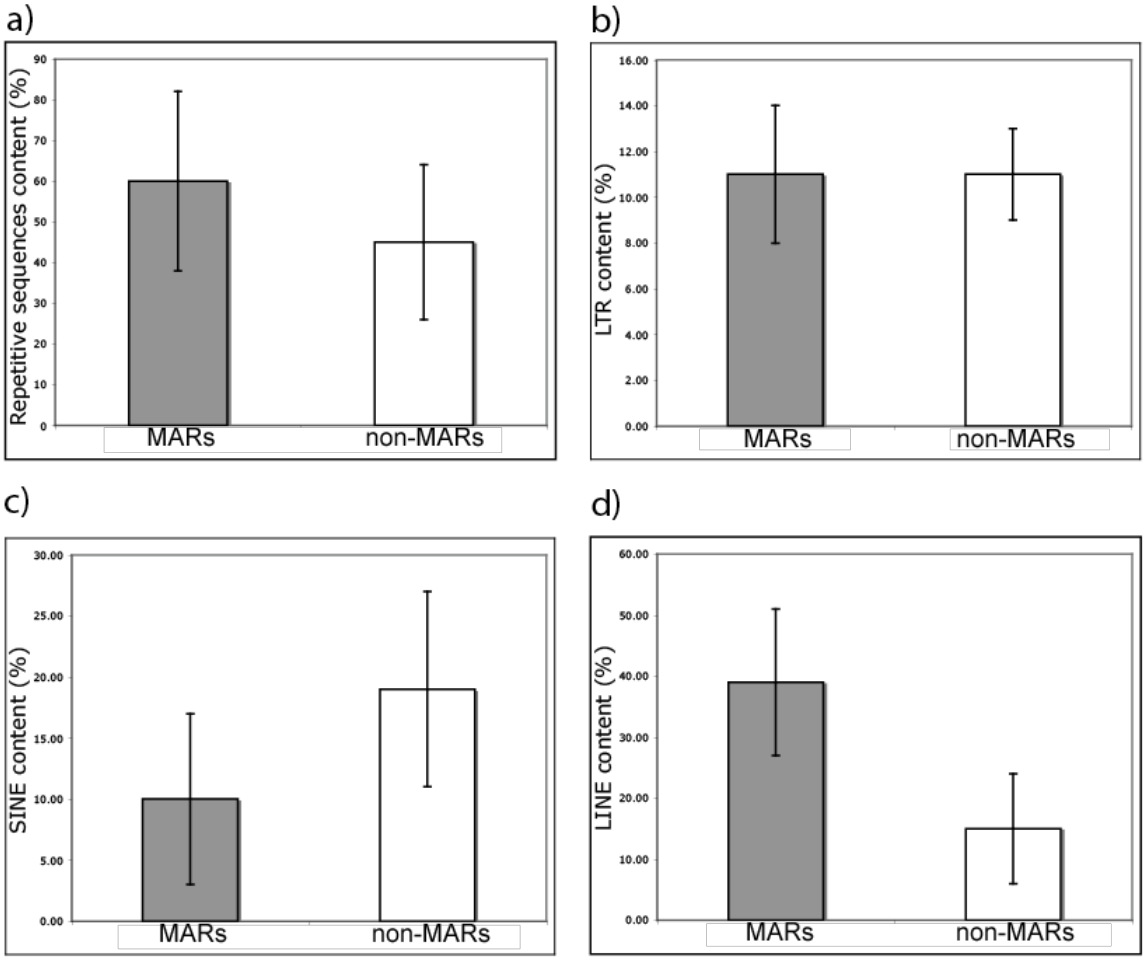

Figure 4.12 Sequence analysis of MARs and non-MARs in relation to different classes of repetitive sequences.

a) The average repetitive content (P-value of 1), **b)** the LTR (P-value of 1), **c)** the SINE (P-value < 0.006) and **d)** the LINE abundance were calculated for MARs and non-MARs (P-value < 0.002). The bars represent the mean values of two biologically independent experiments and show the standard deviation.

4.2.6 Comparison between matrix attachment regions identified experimentally and *in silico*

Several computer programmes have been developed to predict the positions of MARs. Each of these recognises a unique set of motifs associated with MARs. SIDD identifies regions with a tendency for base-unpairing (Benham et al., 1997). MARScan identifies a bipartite sequence common to a large number of eukaryotic MARs (van Drunen et al., 1999). ChrClass identifies sequences that associate with different components of the nuclear matrix including lamins A, B and C (Glazko et al., 2001). Here, the performance of these, SIDD, MARScan and ChrClass, was assessed by comparing *in silico* predictions with the experimental data.

The three programmes predicted different numbers of MARs for the MHC ranging from 187 to 617. The total number of predicted MARs matching (true positives) or not matching (false positives) the experimentally defined MARs in MRC5 fibroblasts is shown in Table 4.4. As expected for such heterogeneous sequences the number of false positives is high. ChrClass predicted most true and false positives while MARScan predicted the fewest.

The accuracy of each computer programme is also expressed using specificity and sensitivity scores (Table 4.4). The specificity is defined by the formula: (number of true positives)/(total number of predictions) while the sensitivity is defined by the formula: (number of true positives)/(total number of expected MARs). The highest sensitivity was shown by ChrClass and the lowest by MARScan. A low specificity was found for all three programmes. Overall the programme with the best performance in terms of sensitivity and specificity was ChrClass.

Of the 117 predicted experimentally identified MARs, 18 were not predicted by any of the three programmes, suggesting that additional features should be considered to develop more accurate MAR prediction programmes.

	ChrClass	SIDD	MARScan
Total Predicted	617	543	187
True Positive	97	63	27
False Positive	520	480	160
False Negative	27	61	97
Specificity	0.16	0.12	0.14
Sensitivity	0.83	0.54	0.23

Table 4.4 Comparison between different MAR predicting algorithms.

True positives are predictions coinciding with the experimentally defined MARs in MRC5 fibroblasts. False positives are predictions that do not coincide with experimentally identified MARs. False negatives are the experimentally identified MARs that are not predicted by each programme. The specificity is given by the total number of true positives divided by total number of predictions. The sensitivity is determined by the number of true positives divided by the total number of experimentally identified MARs.

4.3 Discussion

In this chapter, MARs were identified across the MHC in human MRC5 fibroblasts using MHC tiling path arrays. The key finding is that MARs are non-randomly distributed across the MHC as they cluster at specific locations. Although a unique sequence consensus was not identified, the sequence context of MARs was defined. MARs reside within DNA regions of high AT and repetitive sequences content and low gene density. Most of the MARs identified are situated within intergenic domains of the MHC class I and II regions, closely flanking several genes that are critical for immunity, as well as at the boundaries of the MHC sub-regions. This suggests that the folding of the MHC might contribute to transcriptional regulation by defining distinct chromatin domains.

The 117 MARs identified in the MHC divide the locus into loop domains averaging 35 kb in size. This is similar to the 36 kb loop size estimated for human chromosome band 16q22.1, but smaller than the 50-200 kb predicted by the 'multi-loop sub-compartment model' which appears to be the model most compatible with recent experimental findings (Munkel et al., 1999; Ottaviani et al., 2008b; Shaposhnikov et al., 2007). It is important to note that the frequency of MARs has been shown to vary greatly across the human genome. For example, telomeres have one MAR per kb, while the 19q13.12 band has one MAR every 88 kb (Chernov et al., 2002; Luderus et al., 1996). A substantial variation in MAR density is present between the MHC sub-regions. More than 90% of the MARs identified are in the class I and II regions, which have loop sizes averaging 37 kb and 18 kb respectively. The MHC class III and extended class II regions contain less than 10% of the MARs and have larger loop sizes of 147 kb and 72 kb, respectively.

The class I and II regions are characterised by a higher AT content and repetitive sequences, lower gene density and lower overall conservation compared to the class III and extended class II regions. Computer-assisted analysis was used to define the characteristics of MAR sequences. A positive correlation was found between MARs and LINEs. This finding is in agreement with a study of experimentally identified sequences collected in the MAR transaction database (http://transfac.gbf.de/SMARtDB/) (Jordan et al., 2003). The study revealed that LINEs are over-represented by more than twofold in human MARs relative to their overall

density in the human genome (Jordan et al., 2003). Analysis of the 1 Mb imprinted domain on mouse chromosome 7 also showed a similar LINE enrichment in MAR sequences (Purbowasito et al., 2004). Another positive correlation was found between MARs and high AT content. The AT content of MAR sequences ranged from 39% to 67% with an average of 58%, which is higher than that of surrounding regions, which is 53%. This is compatible with studies in mice and humans, which showed that the AT content of MARs is often only slightly higher than that of their whole genomes (Purbowasito et al., 2004; Shaposhnikov et al., 2007). One of the best-characterised MARs, located in the FRA1B fragile site of CEM leukaemic cells, is almost entirely constituted of AT-rich minisatellites that appear to strengthen its attachment to the nuclear matrix (Jackson et al., 2003). Nevertheless, a large body of evidence suggests that high AT content alone is not sufficient to ensure binding to the nuclear matrix (Bickmore and Oghene, 1996; Das et al., 1993; von Kries, 1991).

In this chapter, the relationship between MAR sequences and other genomic features including gene density, coding or non-coding DNA and evolutionary conservation were examined. There is a negative correlation between MARs and gene density. This relationship is most clearly seen in the classical class III region, which has the highest gene density in the whole genome and the lowest MAR frequency among the MHC sub-regions. Similar results were observed by a recent study analysing MARs in chromosome 16 in AoAf and HeLa cells, suggesting that gene-rich regions are generally less topologically constrained than gene-poor regions (Linnemann et al., 2008). My findings show that MARs are frequently located within intergenic sequences.

Clusters of MARs were mapped near the boundaries separating the MHC sub-regions and flanking a significant number of classical class I and II genes including *HLA-F*, *HLA-G*, *HLA-C* and *HLA-DRA*. This indicates that the interaction between the nuclear matrix and MARs could define chromatin domains, in agreement with several reports suggesting that MARs could function as boundary elements by topologically constraining the DNA into loops or by forming a physical barrier that blocks interactions between promoters and unrelated enhancers (Dunn et al., 2003; Nabirochkin et al., 1998).

Sequence analysis of MAR sequences did not identify a unique MAR consensus and showed a negative correlation between MARs and evolutionary conserved domains.

However, studies in eukaryotic organisms ranging from plants to mammals, clearly show that the tethering function of MARs is conserved across different species (Eivazova et al., 2007; Ioudinkova et al., 2005; Michalowski et al., 1999; Ostermeier et al., 2003). These findings reinforce the view that the interactions of MARs with the nuclear matrix is likely to depend on DNA sequence contexts generating distinct structures rather than on a primary sequence consensus.

The positions of the experimentally identified MARs in the MHC were compared to those predicted by three different MAR prediction programs, namely ChrClass, SIDD and MARScan. The general conclusion is that while ChrClass showed high sensitivity, outperforming SIDD and MARScan, and might be considered as valuable tool for genome-wide studies, each of these computer programs showed reduced specificity. Hence, the identification of an increasing number of MAR sequences will be essential to develop more accurate MAR prediction algorithms.

The MHC tiling path array was used successfully to identify MARs across the entire MHC locus. However, a disadvantage is that MARs are 100-2000 bp and the 2 kb resolution of the arrays is not sufficient to define their sequence at the base pair level. A limitation of the data analysis strategy is that the stringent threshold used to define MARs might have led to an under-representation of the number of MARs across the MHC. The threshold of array enrichment to define MARs was set at approximately 4 standard deviations of the mean on the basis of real-time PCR reactions that were carried out to validate the array data. Values of enrichment below such a threshold could not be consistently validated with real-time PCR. One reasonable assumption might be that the low enrichment of several clones of the MHC tiling path arrays could represent background signal due to hybridisation of non-specific DNA sequences. Using the relatively stringent threshold of 4 standard deviations, the rate of false positives was estimated from the real-time PCR results to approach 0% but the rate of false negatives was predicted to be approximately 2%.

Not all previous investigations agree with our findings. For example, the study by Shaposhnikov et al. (2007) showed that 98% of the MARs identified in band 16q22.1 in HeLa S3 cells are localised within genes (Shaposhnikov et al., 2007). This contrasts with our findings indicating that more than 80% of the MARs in the MHC are situated outside genes. The same group mapped MARs within the 19q13.12 region and found

that only approximately half of the MARs were situated within genes (Chernov et al., 2002). Here, the differences in MAR distribution observed in several loci are probably not due to the different methodologies used to identify MARs but are likely to reflect the distinct spatial organisation of different regions of the genome. Given the heterogeneity of the human genome it is reasonable to expect the data to differ substantially.

Inconsistencies between different studies could also due to the cell or tissue types used, or due to the replication and transcription status of a given genomic domain (Heng et al., 2004; Ostermeier et al., 2003). Several studies, including one characterising SATB1, have shown that MARs can bind to cell type specific matrix proteins suggesting that a proportion of MARs are cell type specific (Cai et al., 2006). Other studies have indicated that the binding of certain MARs to the nuclear matrix correlates with transcription or replication of the loci with which they are associated (Anachkova et al., 2005; Eivazova et al., 2007; Heng et al., 2004). The type of procedure used to isolate MARs might also be responsible for inconsistencies. A comparison of different matrix extraction procedures showed that 2 M NaCl led to the isolation of several MARs that differed from those isolated using 25 mM LIS and 0.65 M ammonium sulphate (Donev, 2000).

In conclusion, I have identified MARs across the MHC region of human MRC5 fibroblasts. This is the first study to map MARs within a 4 Mb locus at a relatively high resolution. The frequency and distribution of the MARs as well as their sequence composition were determined. Such important insights will help to elucidate the relationship between DNA folding and genomic function.

5 Genomic anchors in PGF B lymphoblastoid cells and IFN-γ induced MRC5 fibroblasts

5.1 Introduction

The relationship between nuclear organisation, chromatin architecture and gene expression is not well understood. Several studies indicate that transcription of certain loci can be modulated by their re-positioning within the nucleus and by interactions with loci on different chromosomes (Kioussis, 2005; Reddy et al., 2008; Spilianakis et al., 2005). In addition, certain gene clusters and gene-rich regions have been found to loop out of their CT upon transcriptional up-regulation (Chambeyron and Bickmore, 2004; Volpi et al., 2000; Williams et al., 2002). These studies raise the hypothesis that the movement of chromatin toward distinct nuclear microdomains could promote transcription and facilitate the co-regulation of functionally related genes (Kumaran et al., 2008). In order to test this hypothesis, it is critical to characterise the proteins and the DNA sequences that are involved in the re-organisation of chromatin.

The identification of components of the nuclear matrix such as the SATB1 transcription factor, RNA and DNA polymerases suggests that transcription and replication occur at the nuclear matrix (Ottaviani et al., 2008b). Nuclear myosin, which is found in the nuclear matrix, has been found to mediate long-range chromosome movements at interphase (Chuang et al., 2006). Topoisomerase II, a major component of the nuclear matrix, is able to promote the re-structuring of chromatin (Adachi et al., 1989). The nuclear matrix thus not only provides the structural scaffold where key nuclear processes are organised, but also may be a major player in chromatin movement.

MARs organise the DNA into loops or folds by mediating the interaction of chromatin with the nuclear matrix. The constitutive association of MARs with the nuclear matrix has been proposed to maintain the integrity of higher-order chromosome architecture (Ottaviani et al., 2008b). The more dynamic transient associations are instead dependent on whether a genomic region is being transcribed or replicated (Heng et al., 2004). Since only a few MARs have been characterised, their role in genomic function and in the context of large-scale chromatin organisation gene expression has yet to be

established. Work on the murine Th2 cytokine locus has revealed that activation of gene expression is accompanied by a recruitment of MARs to the nuclear matrix (Cai et al., 2006). Another investigation showed that the recruitment of MARs to the nuclear matrix was required for gene repression (Alvarez et al., 2000). Additional small-scale studies have suggested that MARs can serve as landing platforms for a wide range of regulatory proteins including chromatin remodelling enzymes, splicing factors, components of the transcriptional machinery and insulators (Cai et al., 2003; Donev et al., 2002; Dunn et al., 2003; Kim et al., 2007; Rajaiya et al., 2006).

In order to gain a more comprehensive understanding of the DNA sequences regulating higher-order chromatin structure upon transcriptional regulation, MARs were mapped across the MHC in cells exhibiting different levels of expression and large-scale chromatin organisation. The cell types used for this study were IFN-γ treated fibroblasts and B lymphoblastoid cells, as described previously (Christova et al., 2007; Volpi et al., 2000), and the findings were compared to those in untreated fibroblasts (Chapter 4). Most genes in the MHC class I and III regions are expressed constitutively in all cell types (Boehm et al., 1997). In contrast, several MHC class II genes are constitutively expressed in IFN-γ induced fibroblasts and B lymphoblastoid cells while are repressed in untreated fibroblasts (Boehm et al., 1997; Rohn et al., 1996).

In this chapter, the reconfiguration of MARs upon transcriptional activation is described. The mapping and comparison of MARs in the different cell types reveals that the uneven distribution of MARs across the MHC is generally maintained from cell to cell. MARs are predominantly located in intergenic regions and at or in close proximity to the MHC class boundaries, thus subdividing the locus into physical domains. However, induction of transcription and extra-chromosomal looping are accompanied by a major recruitment of MARs to the nuclear matrix. This recruitment is mainly adjacent to differentially expressed genes when they are activated, suggesting that the rearrangement of chromatin architecture is closely linked with the regulation of gene expression.

5.2 Results

5.2.1 Biochemical isolation of matrix attachment regions from PGF and IFN-γ induced MRC5 cells

The biochemical isolation of MARs in PGF and IFN-γ induced MRC5 cells was carried out as described in Chapter 4. The optimal incubation time with 2 M NaCl for IFN-γ induced MRC5 nuclei was found to be 12 min, as for uninduced MRC5 nuclei, while it was 9 min for PGF nuclei (Fig 5.1). In each cell type, optimal incubation times produced the largest circular halo structures while incubation times longer than optimal produced halo structures of highly variable shapes.

As for uninduced MRC5 cells, the biochemical isolation of MARs was performed 5 times for each cell line to obtain 5 independent biological replicates. Following separation of the MAR and halo fractions using restriction enzymes, more than 70% of the DNA was in the halo fraction of all samples (Table 5.1).

The MAR controls, HS3 and rep30 and the non-MAR controls, PRM2 and F were interrogated by real-time PCR using the isolated MAR and halo fractions. In both cell types, HS3 and rep30 were found to be enriched in the MAR fraction while PRM2 and F were enriched in the halo fraction (Fig 5.2). In IFN-γ induced MRC5 cells, the MAR fractions contained 76% and 78% respectively of the HS3 and rep30 amplicons, while the halo fractions contained 82% and 84% respectively, of the PRM2 and F amplicons. In PGF cells, the MAR fractions contained 69% and 73% respectively of the HS3 and rep30 amplicons, while the halo fractions contained 69% and 84% respectively, of the PRM2 and F amplicons. These data show that the MAR and halo fractionations were carried out consistently in all the samples.

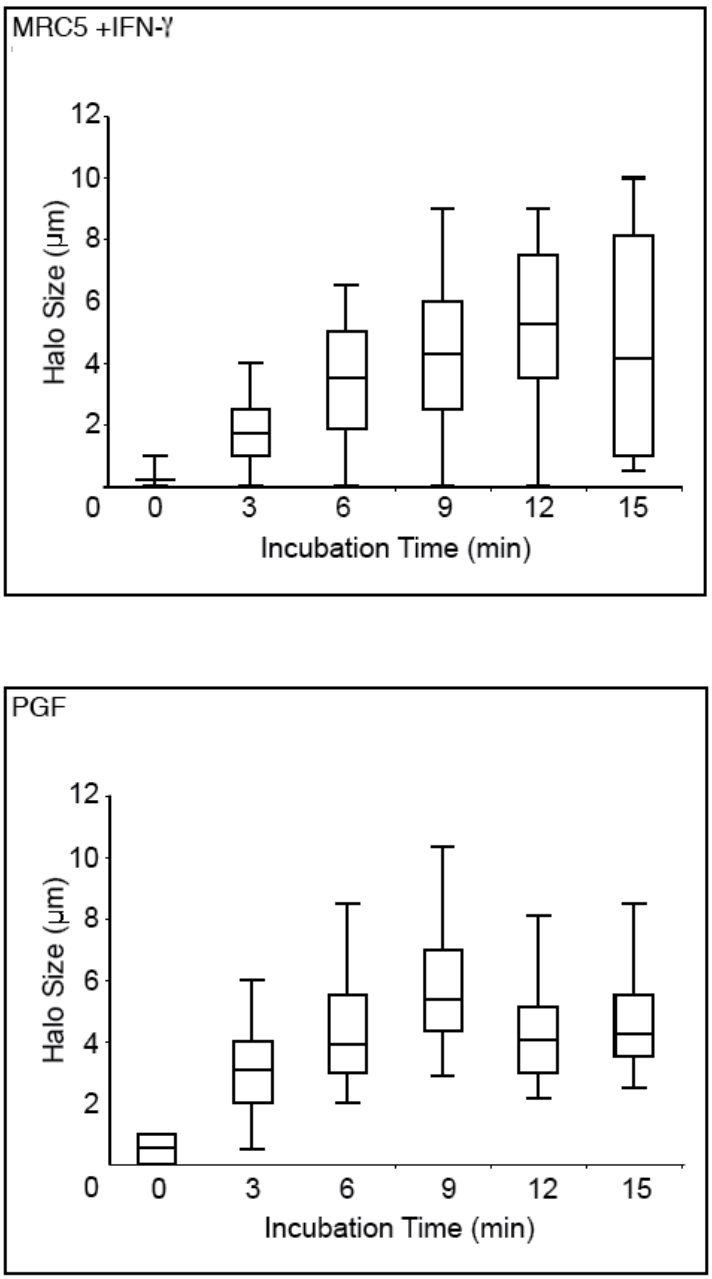

Figure 5.1 Optimisation of halo extraction time in IFN-γ induced MRC5 and PFG nuclei using 2 M NaCl.

The size of halo structures at each time point was determined by averaging the distance of the perimeter of the halo from the edge of the nuclear matrix in 30 nuclei. The optimal incubation time, defined as the minimum incubation time producing the largest halo structures, was found to be 12 min for IFN-γ induced MRC5 nuclei and 9 min for PGF nuclei. Boxplots show the interquartile range of the data, with the horizontal black lines running through each boxplot representing the median values of three independent biological experiments. The whiskers extend as far as 1.5 times the interquartile range.

MRC5 +IFN-γ

Sample	MAR Fraction (%)	Halo Fraction (%)
1	22.6	77.4
2	18.9	81.1
3	25.6	74.4
4	19.8	80.2
5	15.3	84.7

PGF

Sample	MAR Fraction (%)	Halo Fraction (%)
1	16.6	73.4
2	18.4	81.6
3	29.1	70.9
4	21.1	78.9
5	22.3	77.7

Table 5.1 Proportion of DNA in the MAR and halo fractions of IFN-γ induced MRC5 and PFG cells.

The majority of the DNA was consistently found in the halo fraction of IFN-γ induced MRC5 and PGF cells. These proportions were determined by quantifying the amount of DNA in the MAR and halo fractions of each sample and assuming 100% when both fractions were combined.

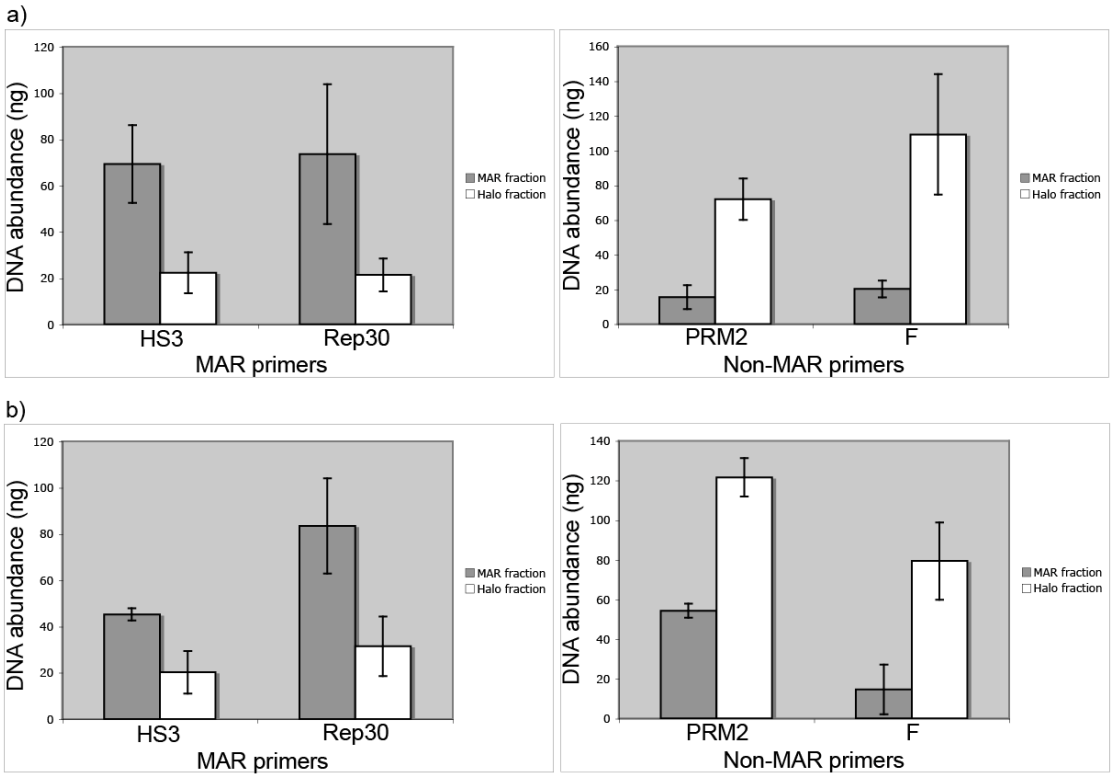

Figure 5.2 Enrichment of MAR and non-MAR amplicons in MAR and halo fractions from IFN-γ induced MRC5 and PGF cells.

Real-time PCR was performed using matrix and halo DNA samples from five biologically independent experiments and primers spanning known MARs (HS3 and rep30) and non-MARs (PRM2 and F). Bars represent the mean of five experiments with standard deviation. The MAR and non-MAR amplicons were enriched in the MAR and halo fractions respectively in samples from **(a)** IFN-γ induced MRC5 and **(b)** PGF cells.

5.2.2 Hybridisation of matrix attachment regions and halo DNA from IFN-γ induced MRC5 and PGF cells onto MHC tiling path arrays

The MARs in the MHC of PGF and IFN-γ induced MRC5 cells were identified using the microarray approach described in the previous chapter. For each cell line two array hybridisations were carried out using samples 1 and 2. The Pearson Correlation Coefficient, which is common measure of correlation, was calculated between all arrays and found to be strongest between duplicates.

The visual examination of the results shows that in PGF and IFN-γ induced MRC5 cells the MARs and halo DNA are unevenly distributed across the MHC (Fig 5.3). This distribution of MAR and halo DNA across the sub-regions of the MHC is remarkably similar to that observed in uninduced MRC5 cells (Chapter 4). In both cell lines the class II regions of the MHC is largely associated with the nuclear matrix, the class I region has long stretches of halo DNA but still contains a significant proportion of MARs while the class III and extended class II regions are mainly constituted of halo DNA.

These results show that, although the overall distribution of halo DNA and MARs across the MHC of the different cell types appears to be conserved, there are important differences. Generally PGF and IFN-γ induced cells have higher enrichments of MARs compared to uninduced MRC5 cells. Furthermore, PGF cells have a higher number of MARs within the class III region of the MHC and a different distribution of MAR and halo DNA within the class I of the MHC compared to IFN-γ induced and uninduced MRC5 cells.

Taken together, the graphical representation of the results from the different cell types indicate that the halo DNA correlates positively with high GC content, gene-dense and highly conserved regions while MARs have high AT content and are located mainly within regions of low conservation and gene-content.

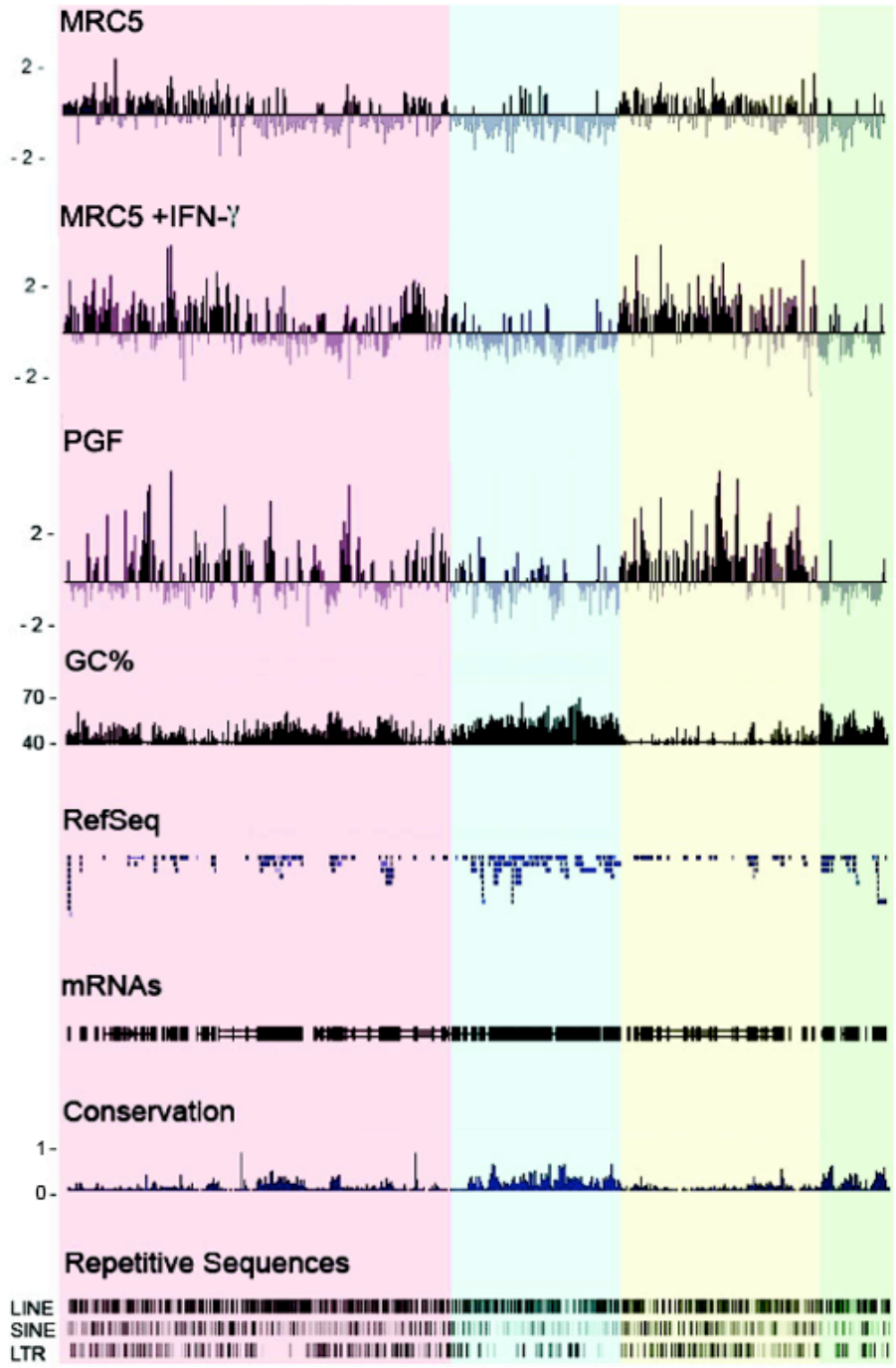

Figure 5.3 Competitive hybridisation of differentially labelled MAR and halo DNA fractions on MHC microarrays using MRC5, IFN-γ induced MRC5 and PGF cells.

Positive values represent MAR-enriched DNA while negative values represent halo-enriched DNA (mean ratios of two independent experiments). GC content, RefSeq gene positions, human mRNAs from GeneBank, regions of consevations between 17 vertebrate species and the repetitive sequences LINE, SINE and LTR across the MHC are also shown.

5.2.3 Quantitative Real-time PCR analysis of matrix attachment regions and halo DNA from IFN-γ induced MRC5 and PGF cells

In order to validate the array results, thirteen regions across the MHC of IFN-γ induced and PGF cells were analysed by quantitative real-time PCR using MAR and halo DNA from samples 3,4 and 5.

In IFN-γ induced MRC5 cells, six regions were enriched in halo DNA while the remaining seven were MAR-enriched. The thirteen amplicons designed to span the selected regions were interrogated using a similar approach to that described in Chapter 4. According to the microarray data, the six regions situated within *HLA-A, HLA-G, HLA-E, HLA-C, HLA-DRB5* and *TAP1* were enriched in the halo DNA fraction. Real-time PCR confirmed high enrichment in the halo DNA fraction for all these regions which had array ratios equal to or lower than -0.128 (Fig 5.4). The seven regions enriched in the MAR fraction based on the microarray data, were situated downstream of *PRRT1, NOTCH4, ZBTB2, HLA-DPB1,* upstream of *C6orf10*, *HLA-B* and *HLA-C* and were validated by real-time PCR (Fig 5.5). These regions had array ratios equal to or higher than 0.437.

In PGF cells, five regions were enriched in halo DNA while the remaining eight were MAR-enriched. The five regions enriched in the halo DNA fraction based on the microarray data were situated in the regions overlapping *HLA-C, HLA-E, HLA-DRB5, TAP1*, and *HLA-G*. Real-time PCR confirmed high enrichment in the halo DNA fraction for all these regions, which had array ratios equal to or lower than –0.144 (Fig 5.6). Of the eight regions enriched in the MAR fraction based on microarray data, six were successfully validated with quantitative real-time PCR (Fig 5.7). These regions located downstream of *PRRT1, NOTCH4, ZBTB2*, upstream of *C6orf10, HLA-B* and *HLA-C* had array ratios equal to or higher than 0.6 ($\log_2$). The remaining two regions upstream of *HLA-A* and downstream of *HLA-DPB1*, which were enriched in the MAR fraction based on array data, could not be validated since they were enriched in the halo fraction by real-time PCR (Fig 5.8). The regions upstream of *HLA-A* and downstream of *HLA-DPB1* had array ratios of 0.319 and 0.127 respectively. This is in agreement with the results in uninduced MRC5 cells indicating that loci with low MAR enrichment using arrays are enriched in the halo fraction with real-time PCR.

Taken together these data confirm that values equal to or higher than 0.4 can be successfully validated with real-time PCR analysis. Thus 0.4 can be used as a threshold to identify MARs across the MHC of uninduced, IFN-γ induced MRC5 and PGF cells.

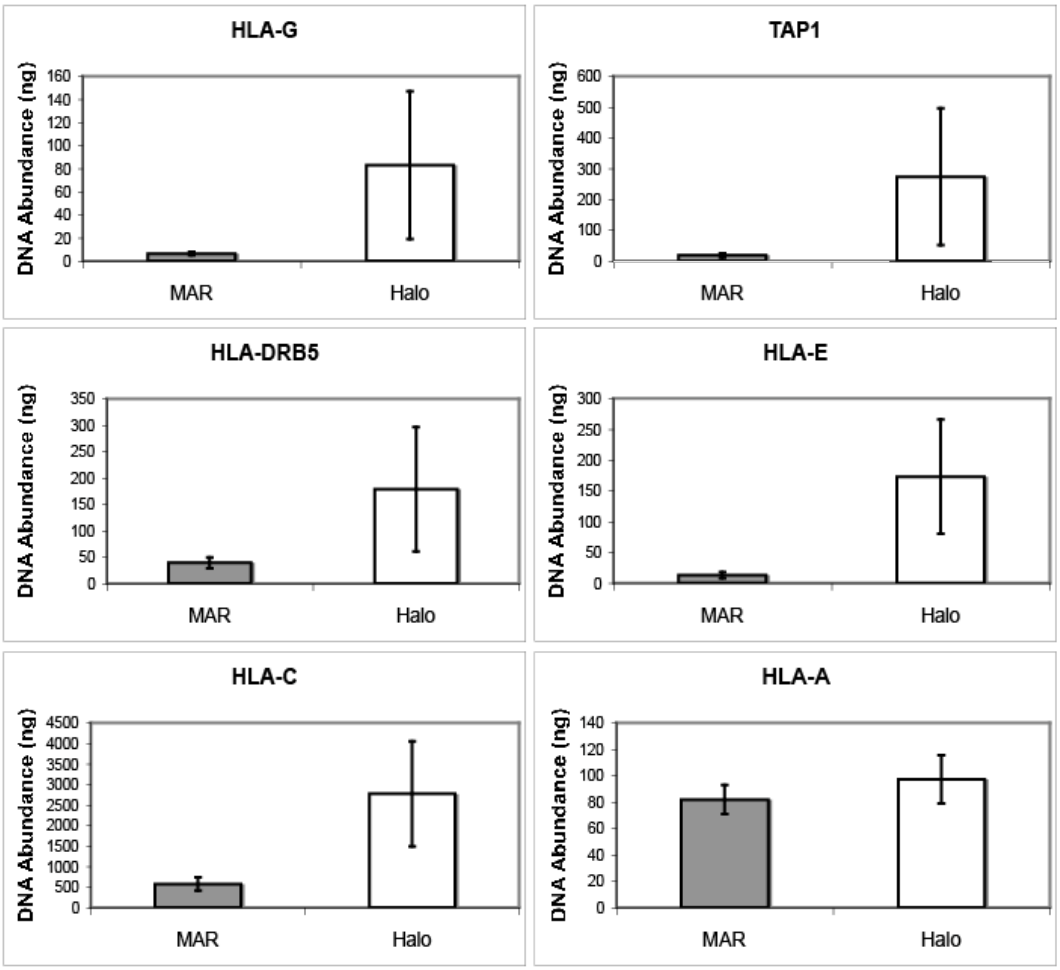

Figure 5.4 Quantitative real-time PCR analysis of halo-enriched loci in IFN-γ induced MRC5 cells.

Real-time PCR was carried out with primer pairs spanning *HLA-G, HLA-DRB5, HLA-C, TAP1, HLA-E* and *HLA-A*, which were enriched in the halo fraction in the array experiments. These loci were confirmed to be highly enriched in the halo fraction. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

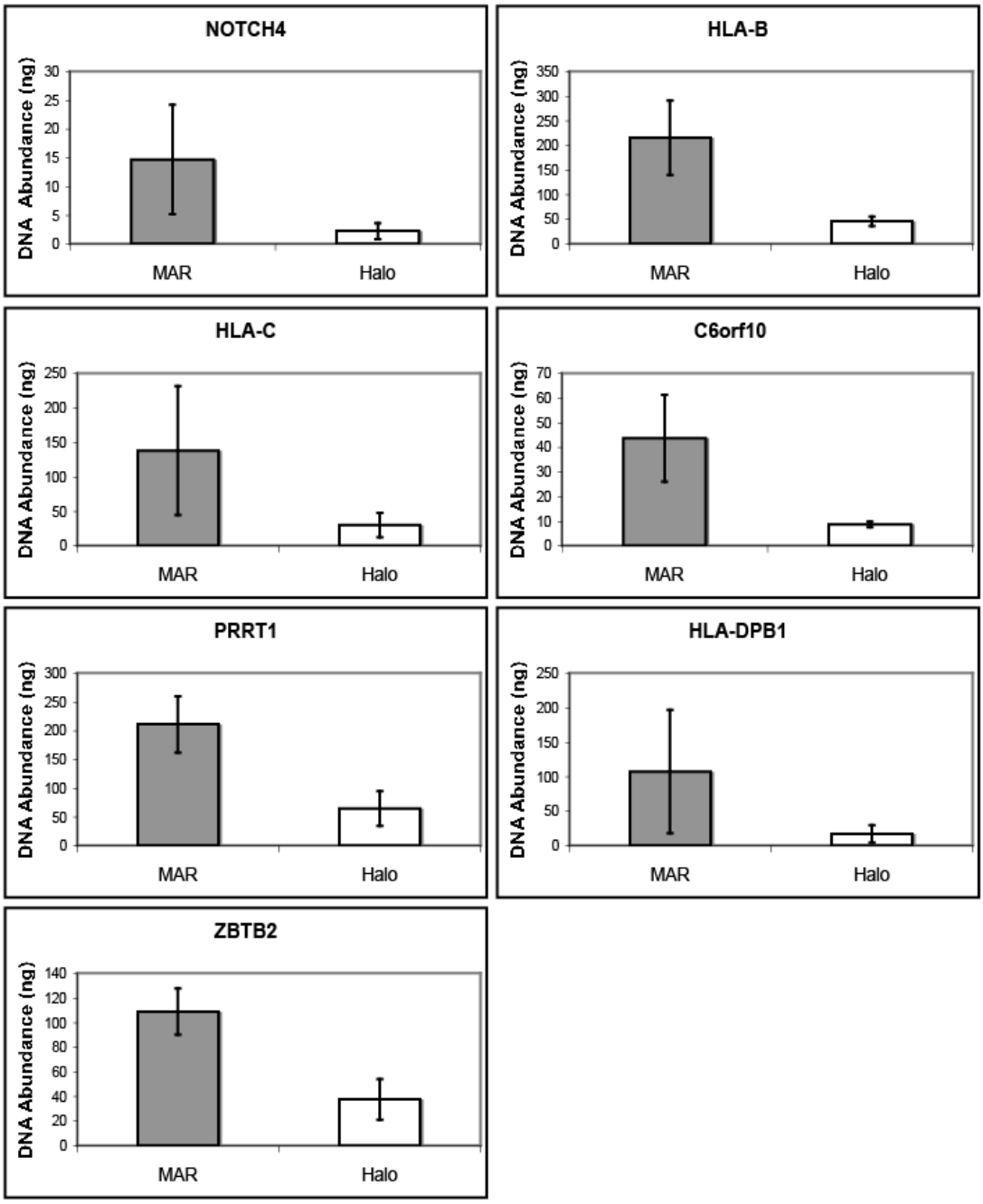

Figure 5.5 Quantitative real-time PCR analysis of MAR-enriched loci in IFN-γ induced MRC5 cells.

Real-time PCR was carried out with primer pairs spanning downstream of *PRRT1, NOTCH4, ZBTB2, HLA-DPB1*, upstream of *C6orf10, HLA-B* and *HLA-C,* which were enriched in the MAR fraction in the array experiments. These loci were confirmed to be highly enriched in the MAR fraction. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

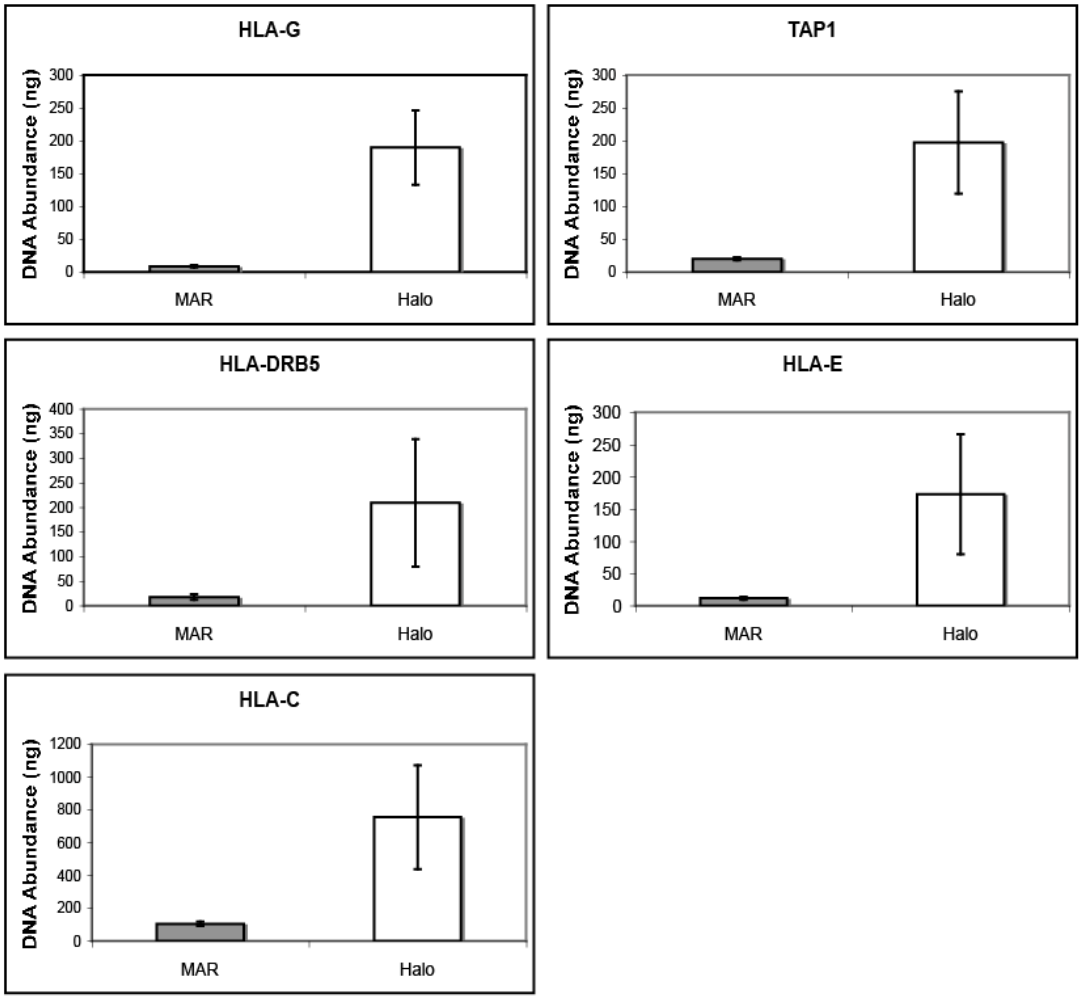

Figure 5.6 Quantitative real-time PCR analysis of halo-enriched loci in PGF cells.

Real-time PCR was carried out with primer pairs spanning *HLA-C, HLA-E, HLA-DRB5, TAP1,* and *HLA-G*, which were enriched in the halo fraction in the results array experiments. These loci were confirmed to be highly enriched in the halo fraction. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

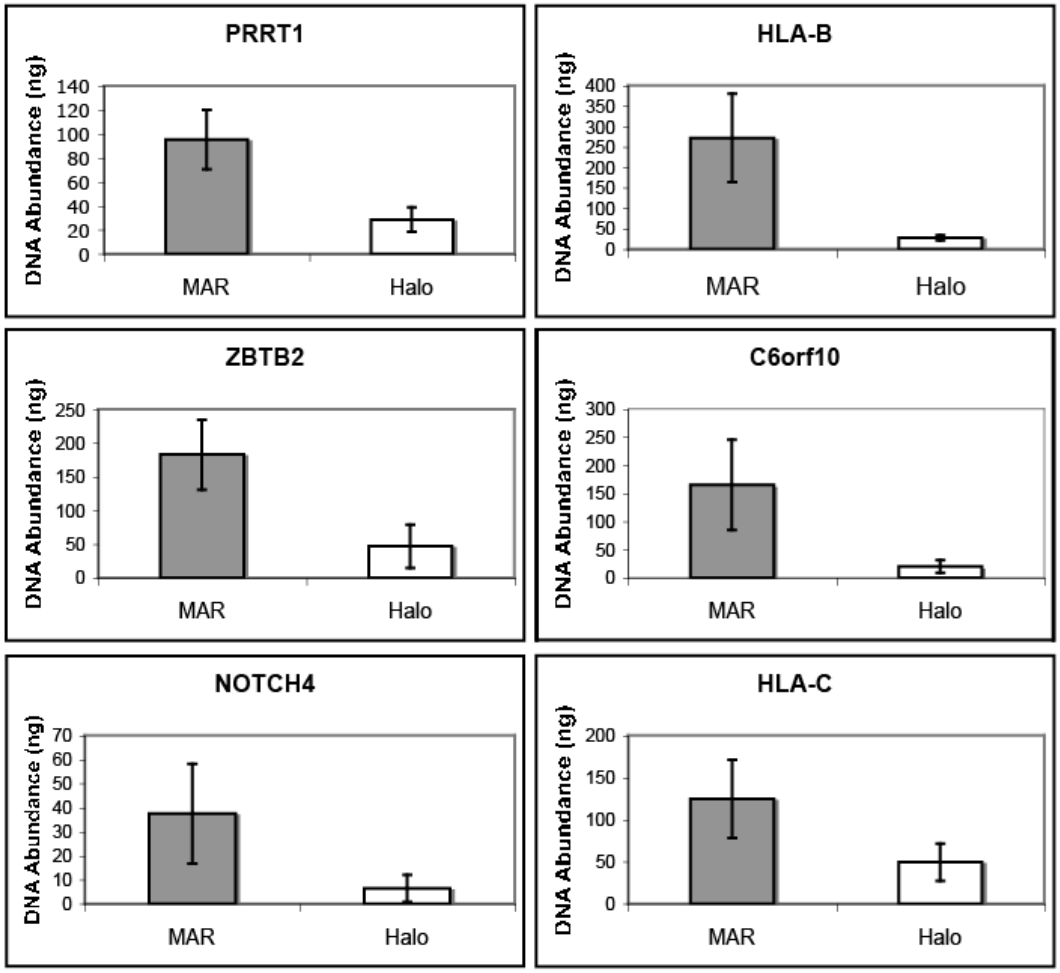

Figure 5.7 Quantitative real-time PCR analysis of loci showing high MAR enrichment in PGF cells.

Real-time PCR was carried out with primer pairs spanning downstream of *PRRT1, NOTCH4, ZBTB2*, upstream of *C6orf10, HLA-B* and *HLA-C,* which showed high enrichment in the MAR fraction in the array experiments. These loci were confirmed to be highly enriched in the MAR fraction. Bars represent the mean of three biologically independent experiments, and show the standard deviation.

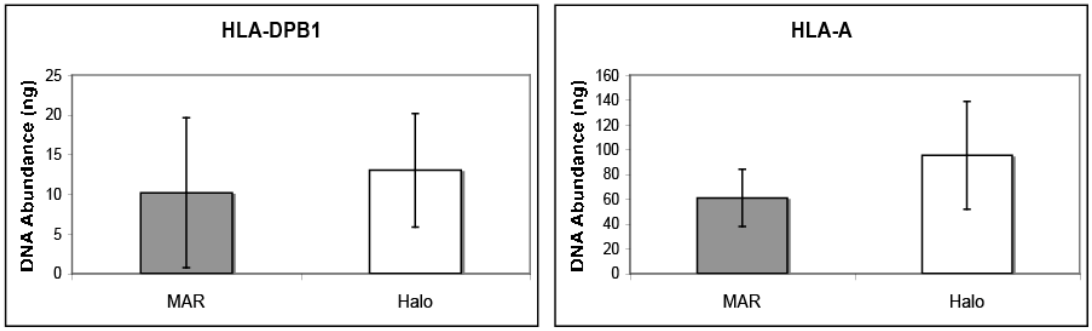

Figure 5.8 Quantitative real-time PCR analysis of loci showing low MAR enrichment in PGF cells.

Real-time PCR was carried out with primer pairs spanning upstream of *HLA-A* and downstream of *HLA-DPB1*, which showed low enrichment in the MAR fraction in the array experiments. Real-time PCR indicated that these regions showed low enrichment in the halo fraction. Bars represent the mean of three experiments, and show the standard deviation.

5.2.4 Position of matrix attachment regions across the MHC of IFN-γ treated MRC5 and PGF cells

Work by Volpi et al. (2000), showed that that giant chromatin loops containing the entire MHC extend outwards from the territory of chromosome 6 with higher frequency in B cells and IFN-γ induced fibroblasts induced than in uninduced fibroblasts. In that study AHB cells were used to represent B cells. Here we confirmed that large-scale looping of the MHC is also observed with high frequency in PGF B lymphoblastoid cells (Fig 5.9).

To determine whether the change in large-scale chromatin organisation and expression of MHC genes are accompanied by a rearrangement of genomic anchors, we identified MARs across the MHC in IFN-γ treated fibroblast and in PGF B lymphoblastoid cells using the MHC tiling path array as previously described in (Fig 5.10). An interactive map of MARs can be obtained by uploading tables in Appendices II-IV into the UCSC Human Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), May 2004 assembly and a detailed profile of MARs is shown in Fig 5.11.

The overall distribution of MARs in the MHC is highly conserved in fibroblast and B lymphoblastoid cells. In all cells studied, over 90% of the MARs identified are located within the classical class I and class II regions while the remaining 10% of MARs are contained in the class III and extended class II regions of the MHC.

There are, however, important differences between the MAR profiles of the different cell lines. A striking increase in the number of MARs was found in IFN-γ induced MRC5 and PGF cells compared to untreated fibroblasts. This increase in the number of MARs is mainly concentrated within the classical class I and class II regions.

In IFN-γ treated fibroblasts and B lymphoblastoid cells, MARs were identified on average every 17 and 15 kb respectively, which is approximately twice the frequency of that in untreated fibroblasts. In IFN-γ induced MRC5 cells, 221 MARs were identified and were subdivided as follows: 111 in the class I, 8 in the class III, 96 in the class II and 6 in the extended class II region of the MHC. The class I, III, II and extended class II regions have one MAR every 16, 120, 9, and 52 kb respectively. In PGF cells, 261 MARs were identified and were subdivided as follows: 115 in the class I, 20 in the class

III, 121 in the class II and 5 in the extended class II region of the MHC. The class I, III, II and extended class II regions have one MAR every 16, 40, 7 and 72 kb respectively.

Of the 221 MARs detected in the MHC of IFN-γ treated fibroblasts, 192 are intergenic whereas 29 contain intragenic regions. In B lymphoblastoid cells of the 261 MARs identified, 205 are located intergenically while 56 contain intragenic DNA.

As observed in uninduced MRC5 cells, in IFN-γ induced MRC5 and PGF cells MARs have been identified at the boundaries separating class I from class III , class III from the class II region and class II from the extended class II region (Fig 5.12) as well as in regions flanking genes such as *HLA-A, HLA-C, HLA-G, HLA-F, HLA-B* and *HLA-DQB1*.

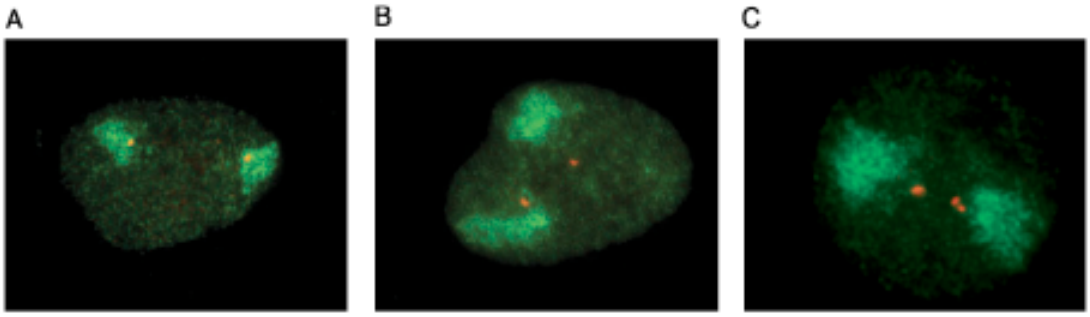

Figure 5.9 Cytological visualisation of giant external chromatin loops carrying the MHC region in relation to the chromosome 6 territory.

(a) FISH image showing the *HLA-DRA* gene (red) located within the chromosome 6 territory (green). **(b)** Only a low basal level of external chromatin loops carrying the MHC of 10% (N = 313) was present in control MRC5 cells. IFN-γ induction of external chromatin loops in MRC5 cells, frequency of giant loop formation increased to 27% (N = 325). **(c)** PGF cells show a high basal level of external chromatin loops, 39% (N = 344). N - number of CTs analysed.

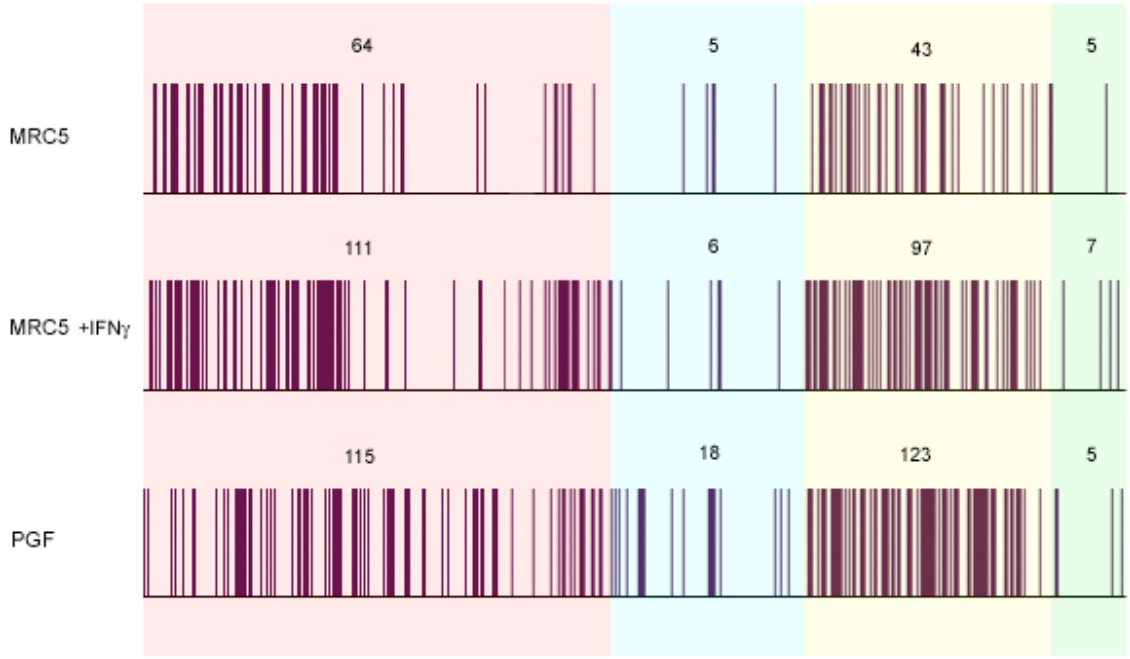

Figure 5.10 MARs in the MHC of MRC5, IFN-γ induced MRC5 and PGF cells.

The mean value for two array hybridisations was calculated for each fragment represented in the MHC tiling path array. Regions enriched in the MAR fraction with mean ratios equal to or higher than 0.4 (log_2) were plotted on the UCSC genome browser (http://genome.ucsc.edu/) and represented on the top row by black vertical lines. The number of MARs in each region is shown above the data plots.

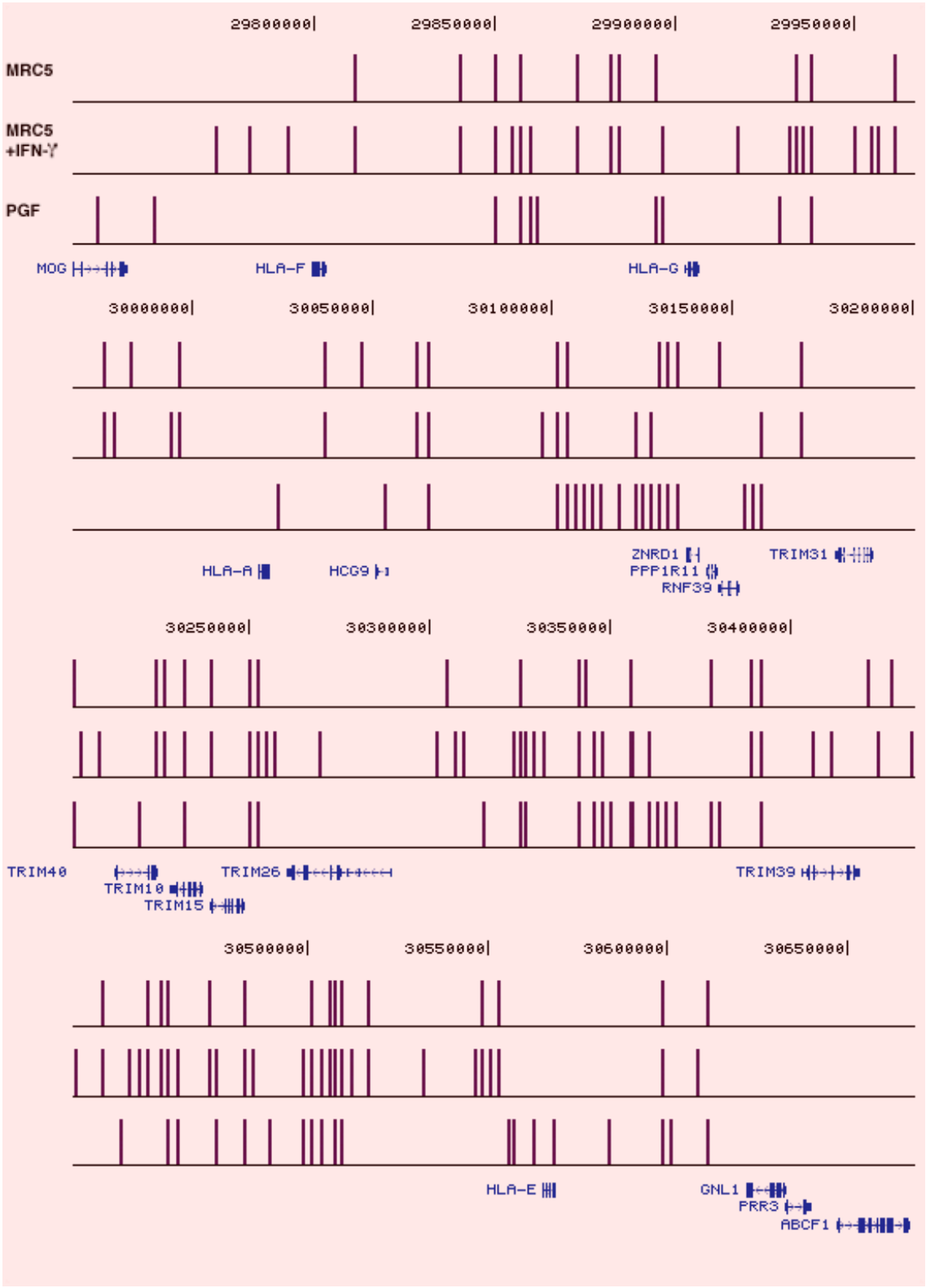

Figure 5.11 (5 pages) Detailed profile of MARs across the MHC of MRC5, IFN-γ induced MRC5 and PGF cells.

MARs are shown relative to Refseq genes from the UCSC web browser (http://genome.ucsc.edu/) in MRC5, IFN-γ induced MRC5 and PGF cells.

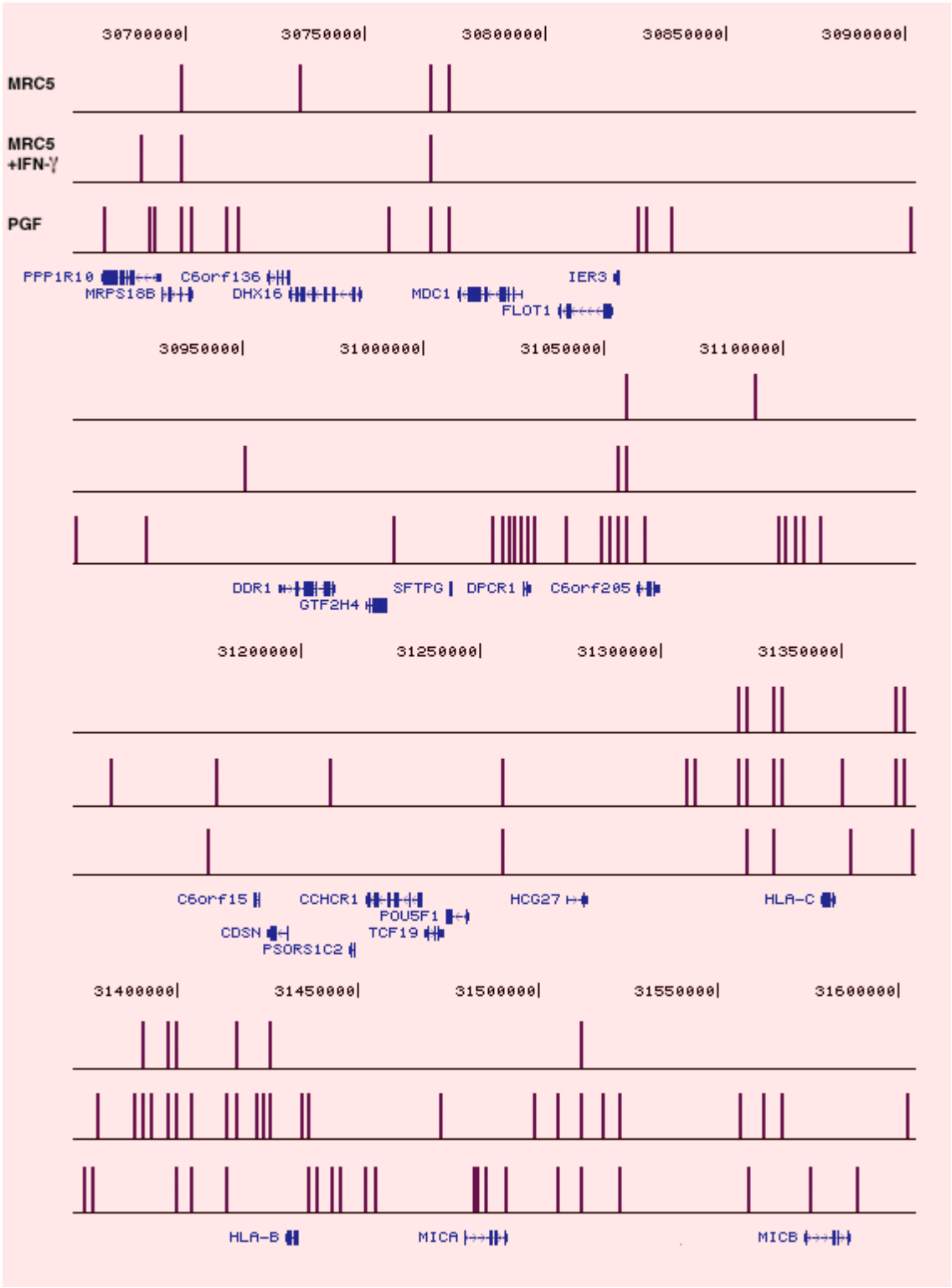

Figure 5.11 (continued).

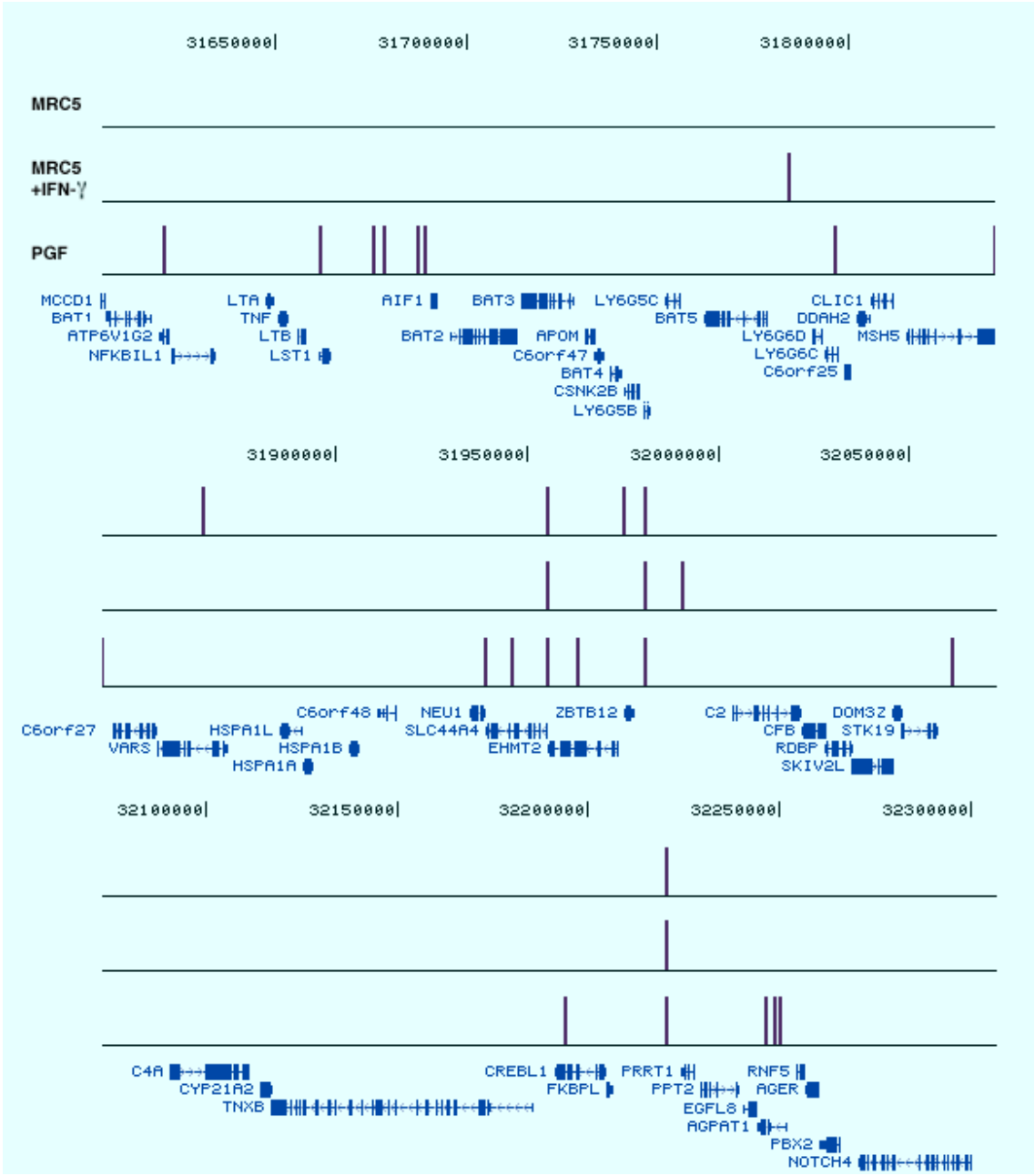

Figure 5.11 (continued).

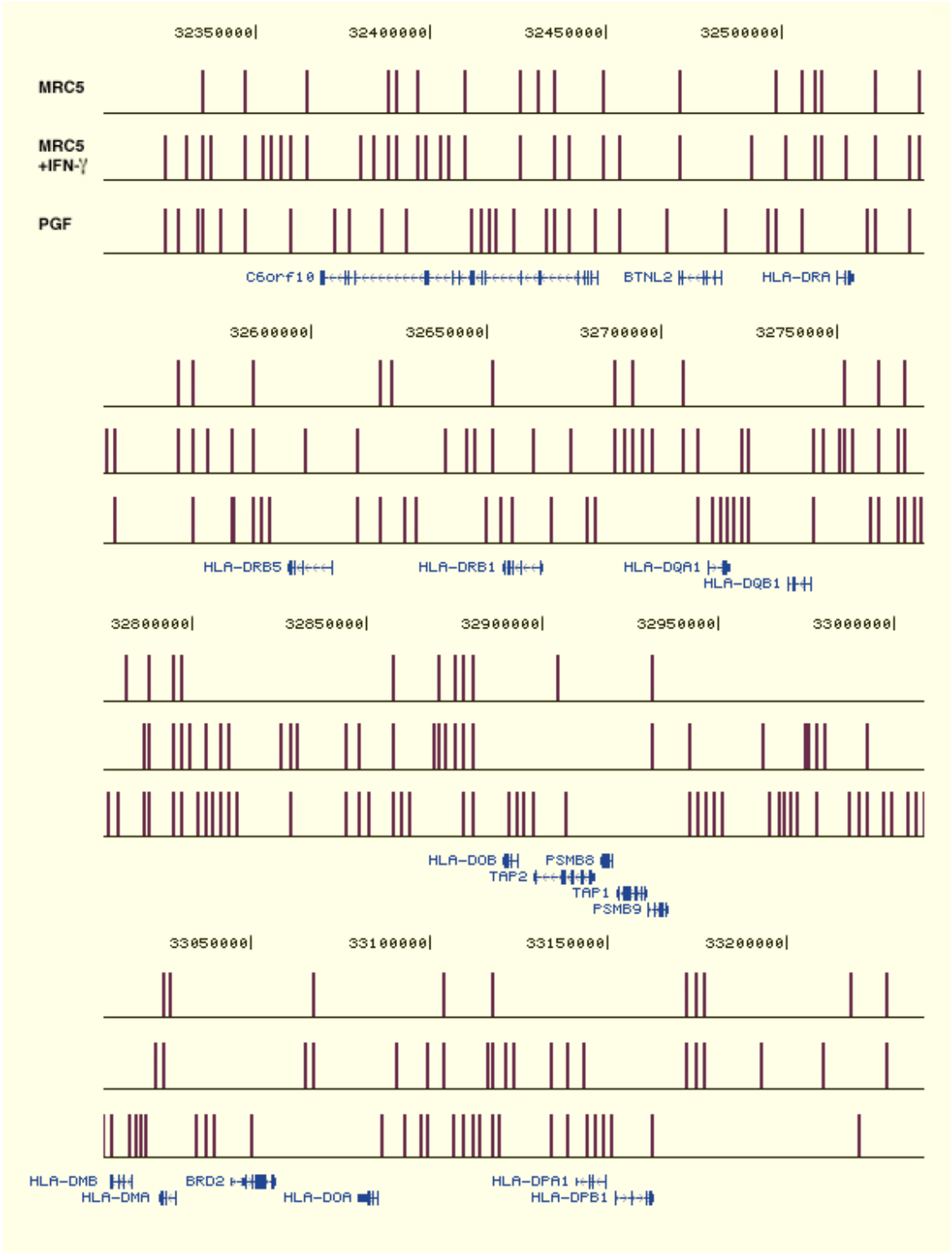

Figure 5.11 (continued).

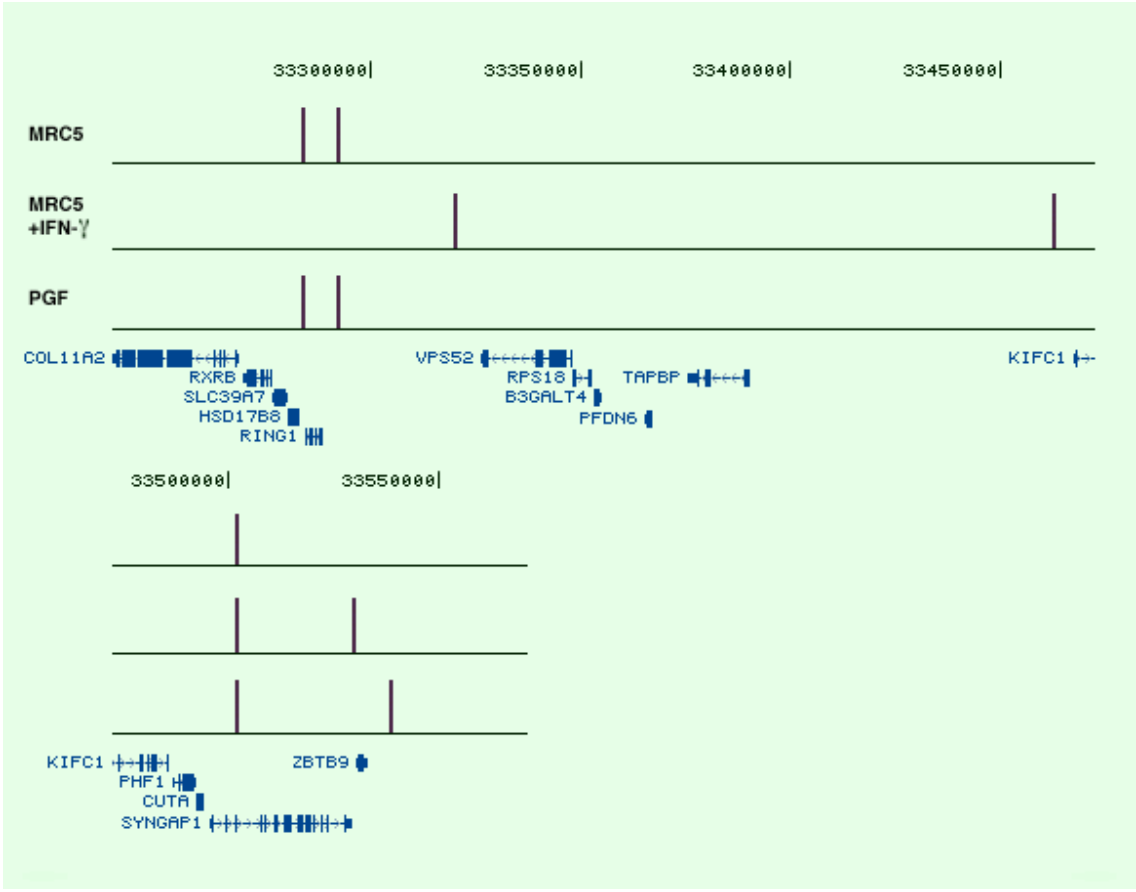

Figure 5.11 (continued).

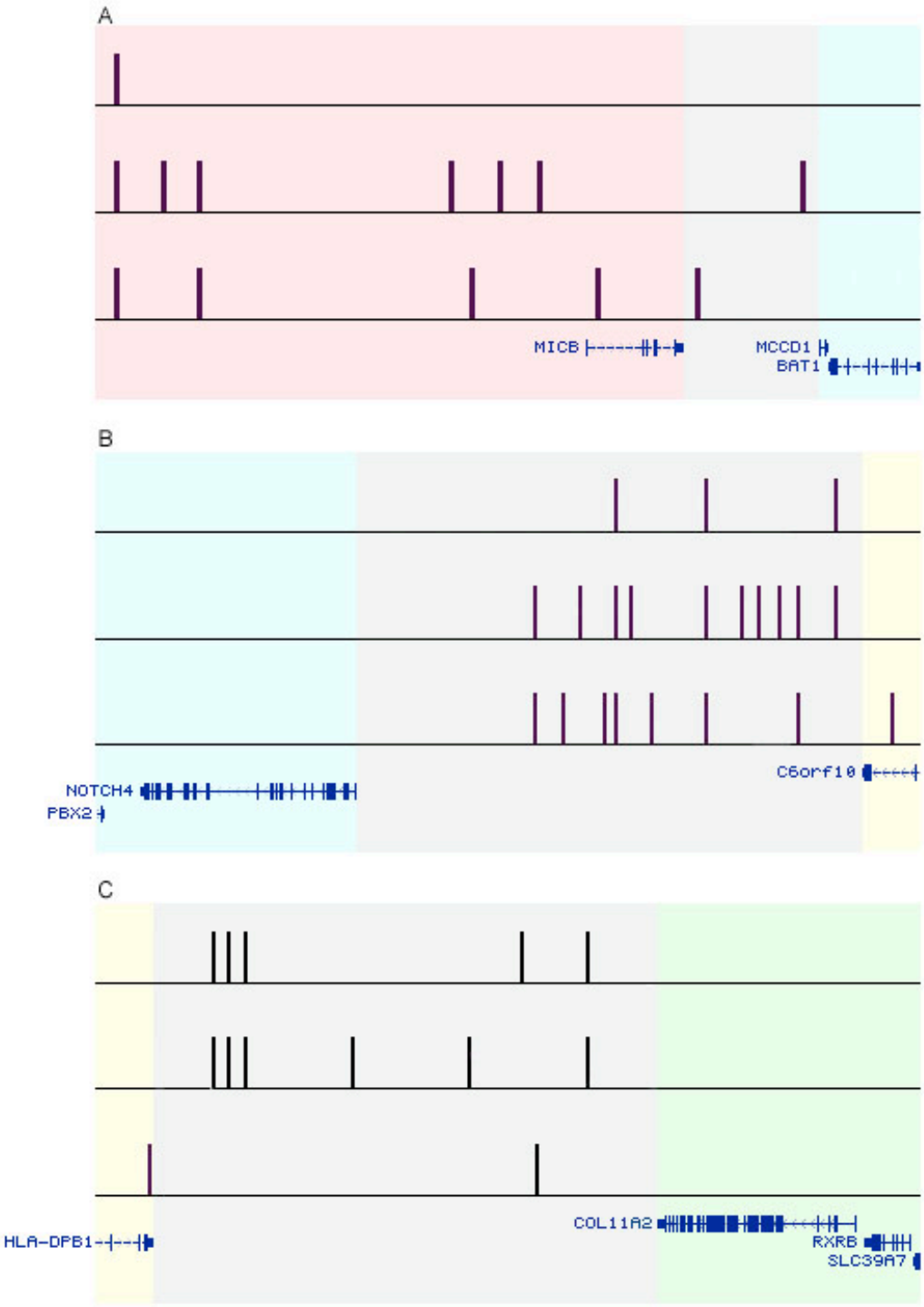

Figure 5.12 Distribution of MARs at MHC class boundaries in MRC5, IFN-γ induced MRC5 and PGF cells.

The boundaries separating **A)** class I from class III, **B)** class III from class II, **C)** class II from the extended class II are highlighted in grey. MRC5, IFN-γ induced MRC5 and PGF cells are shown in the top, middle and bottom row respectively.

5.2.5 Sequence analysis of matrix attachment regions in IFN-γ induced MRC5 and PGF cells

Different genomic properties including conservation, AT content, gene and repetitive sequences density of MARs and non-MARs across the MHC of IFN-γ induced fibroblasts and B lymphoblastoid cells were analysed as described for uninduced fibroblasts.

The average AT content of MARs in IFN-γ induced fibroblasts and B lymphoblastoid cells is 60% and 58% respectively while of halo DNA is 52% and 53% respectively. In general, the MARs identified in IFN-γ induced fibroblasts and B lymphoblastoid cells have a higher AT content than non-MARs (P-value of 0.003 for IFN-γ induced fibroblasts and 0.002 for and B lymphoblastoid cells). However there are a number of MARs that are GC-rich and non-MARs that are AT-rich.

Intragenic sequences constitute more than 40% of non-MARs in both cell types and 16% and 22% of MARs in IFN-γ induced MRC5 and PGF cells respectively (P-value of 0.001 for IFN-γ induced fibroblasts and 0.004 for B lymphoblastoid cells). The Pearson Product Moment Correlation was determined to study the relatiosnhip between gene density and the positions of MARs. The correlation (r) values are consistently negative for all cell types (r-value of -0.23 for uninduced MRC5, -0.32 for IFN-γ induced MRC5 and -0.37 for PGF cells) suggesting that MARs are mainly located in gene-poor regions. The relationship between gene density and the positions of MARs can be visualised in figure 5.13.

The assessment of the conservation of MARs and non-MARs shows that non-MARs are generally more conserved than MARs. However the degree of conservation of these sequences varies substantially between the different cell types. The non-MARs in IFN-γ induced MRC5 cells are three times more conserved than MARs (P-value < 0.003) while the non-MARs in PGF cells are less than twice more conserved than MARs (P-value < 0.005).

In IFN-γ induced fibroblasts the average content of repetitive sequences of MARs and non-MARs is 58% and 38% respectively (P value of 0). In PGF cells the average

content of repetitive sequences of MARs and non-MARs is 48% and 43% respectively (P-value < 0.007). In both cell types, LINEs are over-represented by more than twofold among MARs relative to their overall abundance in non-MARs (P-value of 0) while the SINE content is approximately two times higher in MARs compared to non-MARs (P-value < 0.005). MARs and non-MARs of PGF cells do not differ significantly in relation to the LTR content (P-value of 1). This contrasts to IFN-γ induced MRC5 cells, in which LTRs are over-represented by approximately three-fold in MARs relative to their overall abundance in non-MARs (P-value < 0.005).

The sequence properties of MARs and non-MARs in IFN-γ induced MRC5 and PGF cells are similar to those in untreated MRC5 cells. In summary, compared with non-MARs, MARs are generally AT-rich, enriched in repetitive sequences, depleted of intragenic sequences and show a lower degree of conservation.

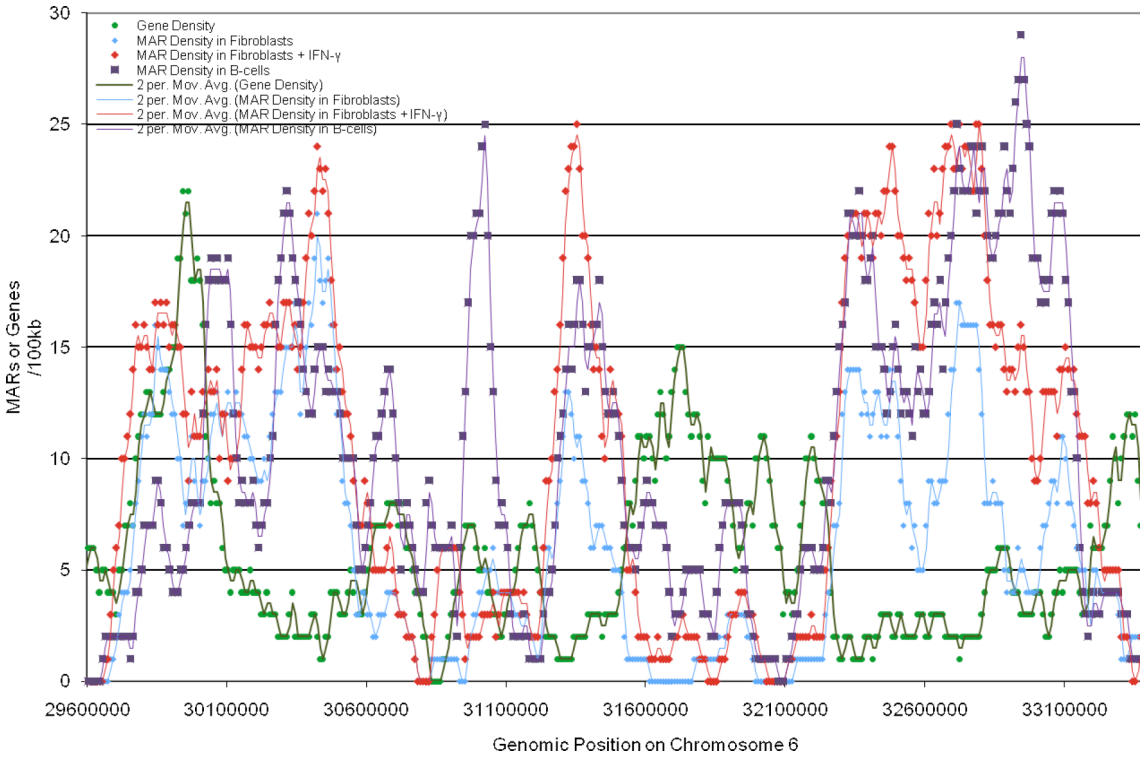

Figure 5.13 Correlation of MARs and gene density in MRC5, IFN-γ induced MRC5 and PGF cells.

Gene and MAR density were binned in 100 kb windows that were slid by 10 kb for each point. The trendlines were plotted based on two points moving average. In general, the different cell types show a negative correlation between MAR and gene density in all cell types.

5.2.6 Classification of matrix attachment regions

To gain a better insight into the roles of MARs in the MHC, we compared their positions in the three different cell types. This comparative analysis identified three main categories of MARs: structural, cell specific and transcription-dependent. The positions of the MARs in each category are shown in tables in Appendices V-X.

MARs constitutively attached to the nuclear matrix in all cell types account for 17% of all MARs (Fig 5.14a). These anchors are defined as structural MARs since their tethering function is preserved regardless of the cell type and transcriptional status of the MHC locus. The majority of the structural MARs identified are situated within the class I region of the MHC.

Approximately 45% of MARs are cell type specific. They are further classified into fibroblast specific, which include MARs present in both untreated and treated fibroblasts (Fig 5.14b), and B lymphoblastoid specific (Fig 5.14c). Most cell type specific MARs are equally distributed within the class I and class II regions of the MHC.

Transcription-dependent MARs are DNA sequences that transiently associate to the nuclear matrix depending on the transcriptional status of the MHC. They include MARs in untreated fibroblasts that are released from the nuclear matrix in IFN-γ treated fibroblasts (Fig 5.14d), MARs recruited to the nuclear matrix in response to IFN-γ treatment (Fig 5.14e) and MARs recruited to the nuclear matrix in both B lymphoblastoid cells and IFN-γ treated fibroblasts (Fig 5.14f). These groups of MARs together constitute 34% of all MARs identified and are predominantly located within the class I and II regions of the MHC.

The remaining 4% of MARs were present in both B lymphoblastoid cells and untreated fibroblasts. To determine whether these MARs are cell specific or transcriptional dependent, it is necessary to characterise MARs across the MHC of other cell lines.

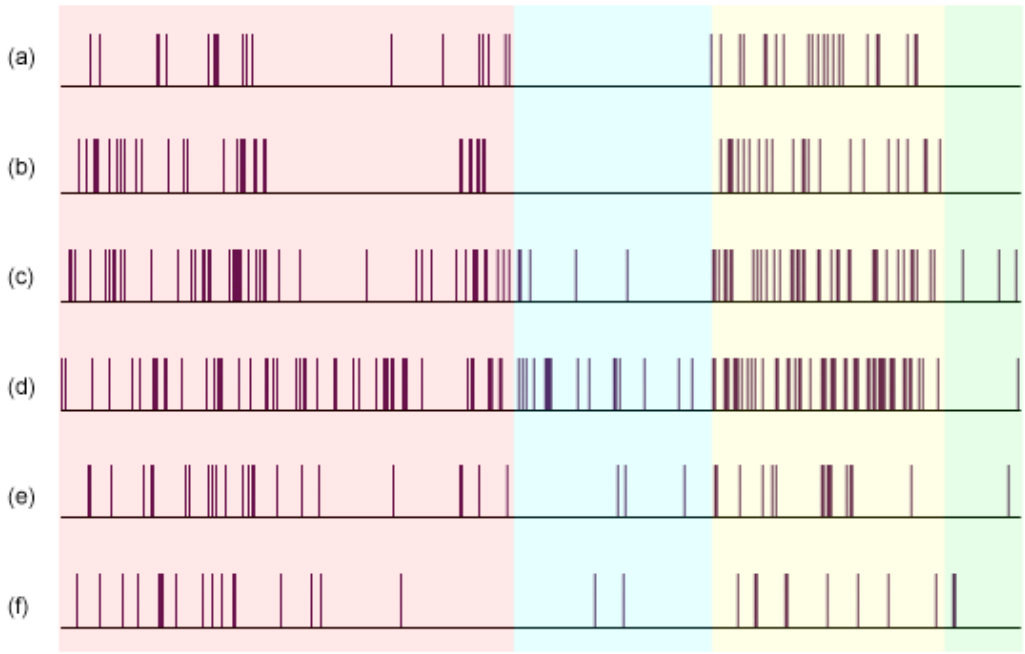

Figure 5.14 Distribution of structural, cell specific and transcription-dependent MARs across the MHC.

Shown are **(a)** structural MARs, **(b)** MRC5-specific MARs, **(c)** PGF-specific MARs, **(d)** MARs of untreated MRC5 cells, **(e)** MARs of IFN-γ treated MRC5 cells and **(f)** MARs of both IFN-γ treated MRC5 and PGF cells.

5.2.7 Distribution of matrix attachment regions relative to differentially and non-differentially expressed genes

The relationship between gene expression and MARs was further investigated by examining the positions of MARs in relation to differentially and non-differentially expressed genes.

Microarrays were used to compare expression of MHC genes of untreated fibroblasts with IFN-γ treated fibroblasts and B lymphoblastoid cells (Fig 5.15). A significant proportion of class II genes and a few class I genes are up-regulated genes in IFN-γ treated fibroblasts and B lymphoblastoid cells. The expression of the class III and extended class II genes in fibroblasts does not differ following IFN-γ treatment, while a proportion of class III genes are up-regulated in B lymphoblastoid cells relative to untreated fibroblasts. As expected, most of the differentially expressed genes code for classical class I and class II molecules (Table 5.2).

The distances of MARs from differentially and non-differentially expressed genes in fibroblasts and B lymphoblastoid cells were calculated to determine how changes in gene expression affect MAR distribution (Fig 5.16). The results show that in both IFN-γ treated fibroblasts and B lymphoblastoid cells the average distance between MARs and differentially expressed genes was over two times shorter than the distance between MARs and non-differentially expressed genes (Fig 5.17). A Mann-Whitney test showed that the distribution was significant with p-values of 0.003 and 0.009, respectively. The average distance between MARs and non-differentially expressed genes in IFN-γ induced fibroblasts is 26 kb, while between MARs and differentially expressed genes, the average distance is 12 kb. This trend is observed for several genes of critical immunological importance including *HLA-C, HLA-DPA1, HLA-DPB1*, and *HLA-DRB1* genes.

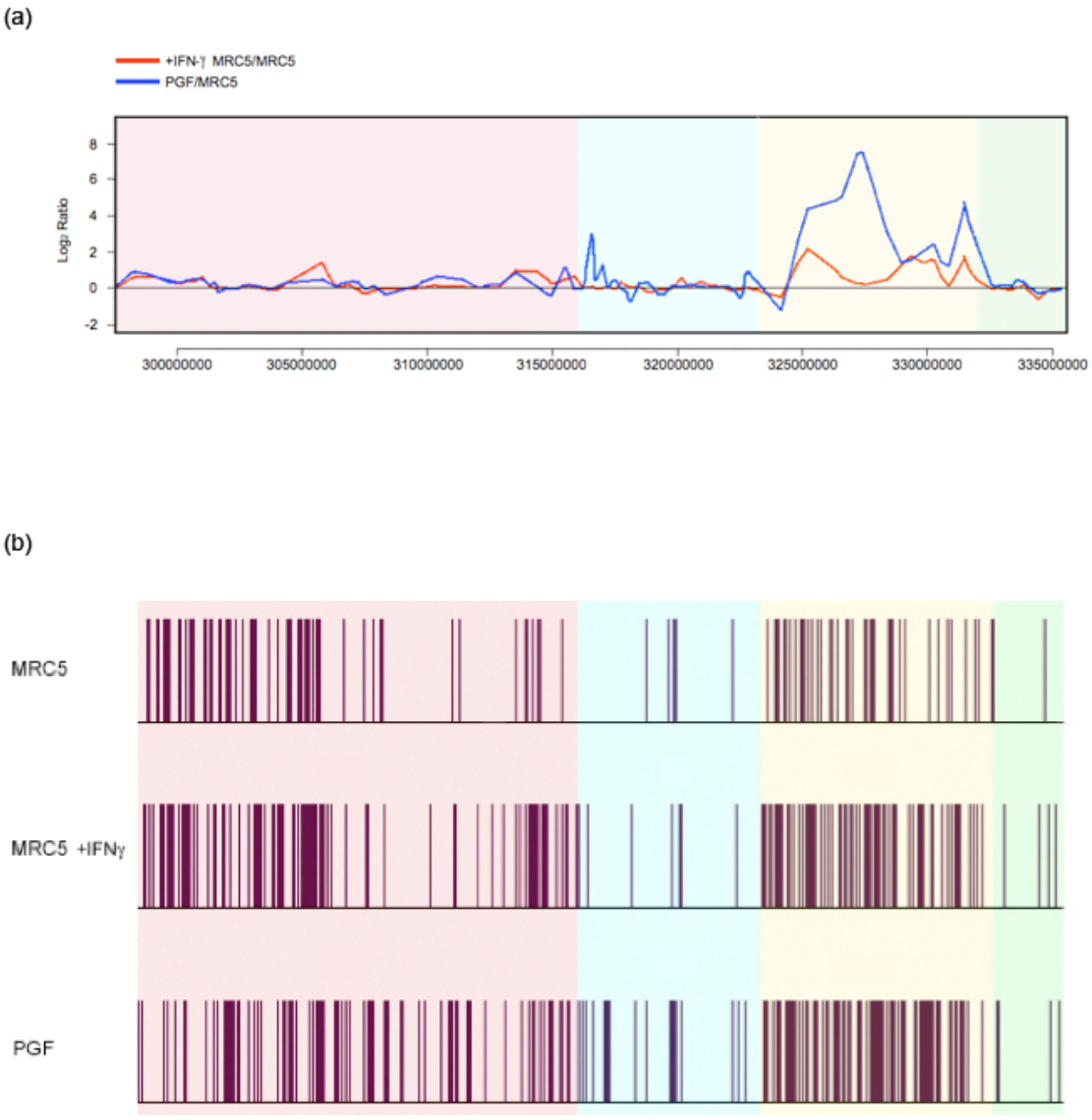

Figure 5.15 Comparison of gene expression levels in the MHC of MRC5, IFN-γ induced MRC5 and PGF cells.

a) The red line represents the $\log_2$ expression ratios of IFN-γ induced MRC5/MRC5 ratios and the blue line represents the ratios of PGF/MRC5. (b) Distribution of MARs in MRC5, IFN- γ treated MRC5 and PGF cells.

	IFN-γ treated MRC5		PGF
Class I	*HLA-F*	**Class I**	*HLA-F*
	HLA-E		*VARS2*
	HLA-C		*HLA-B*
	HLA-B		*HCP5*
	MICB	**Class III**	*LTA*
Class III	*C4*		*TNF*
Class II	*HLA-DRA*		*LTB*
	TAP2		*AIF1*
	PSMB8		*NEU1*
	TAP1		*GPSM3*
	PSMB9	**Class II**	*HLA-DRA*
	HLA-DMB		*HLA-DRB1*
	HLA-DMA		*HLA-DQA1*
	HLA-DPA1		*HLA-DOB*
			PSMB8
			TAP1
			PSMB9
			HLA-DMB
			HLA-DMA
			HLA-DPA1
			HLA-DPB1
		Ext. Class II	*SLC39A7*
			HSD17B8

Table 5.2 Differentially expressed genes in IFN-γ treated MRC5 and PGF cells.

Genes included show at least a two-fold change in up-regulation of expression compared to untreated MRC5 cells.

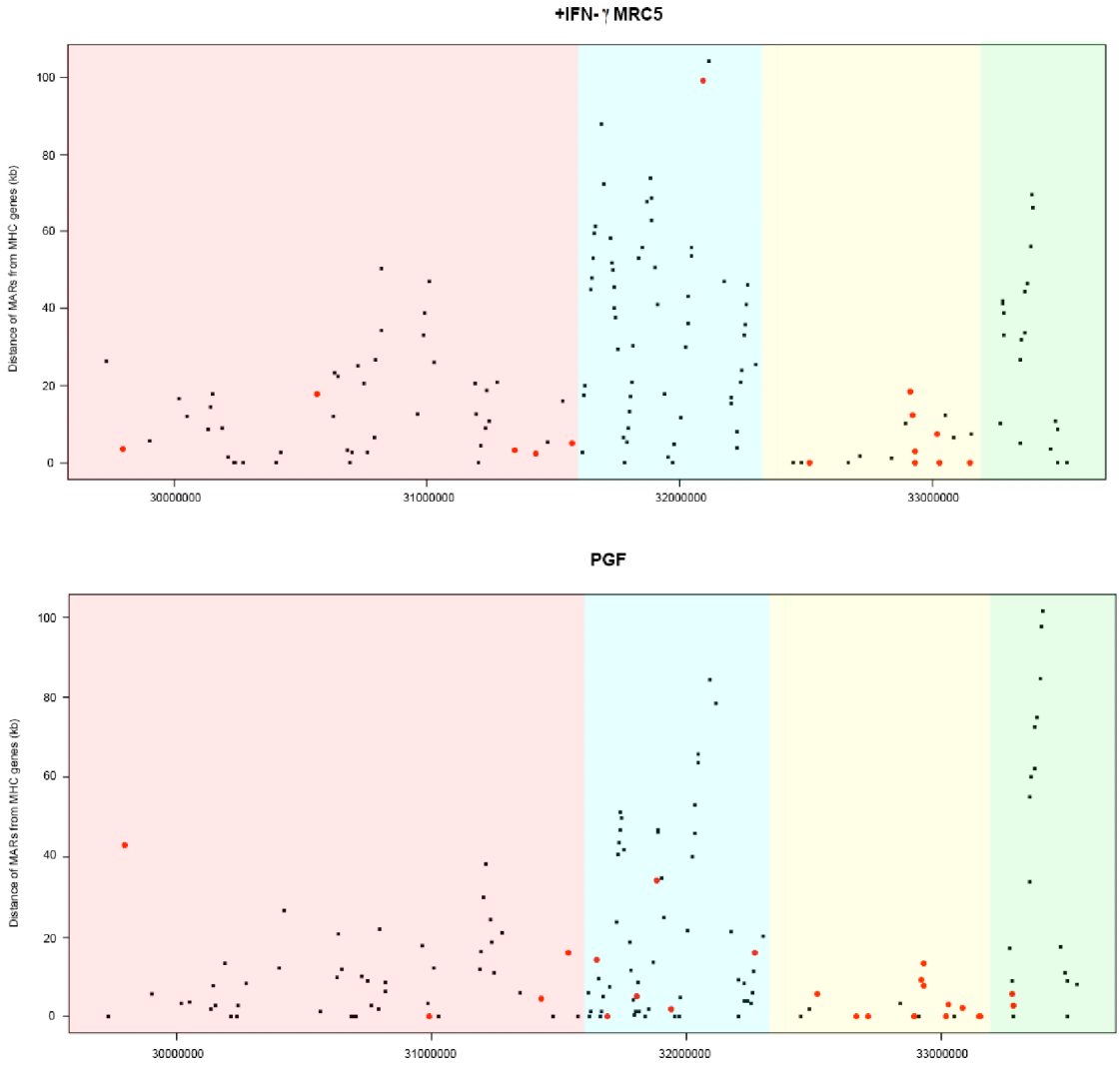

Figure 5.16 Distance of MARs from MHC genes in IFN-γ treated MRC5 and PGF cells.

The genomic position of each gene (x-axis) is plotted against its distance from the closest MAR (y-axis). The black and red dots represent non-differentially and differentially expressed genes respectively.

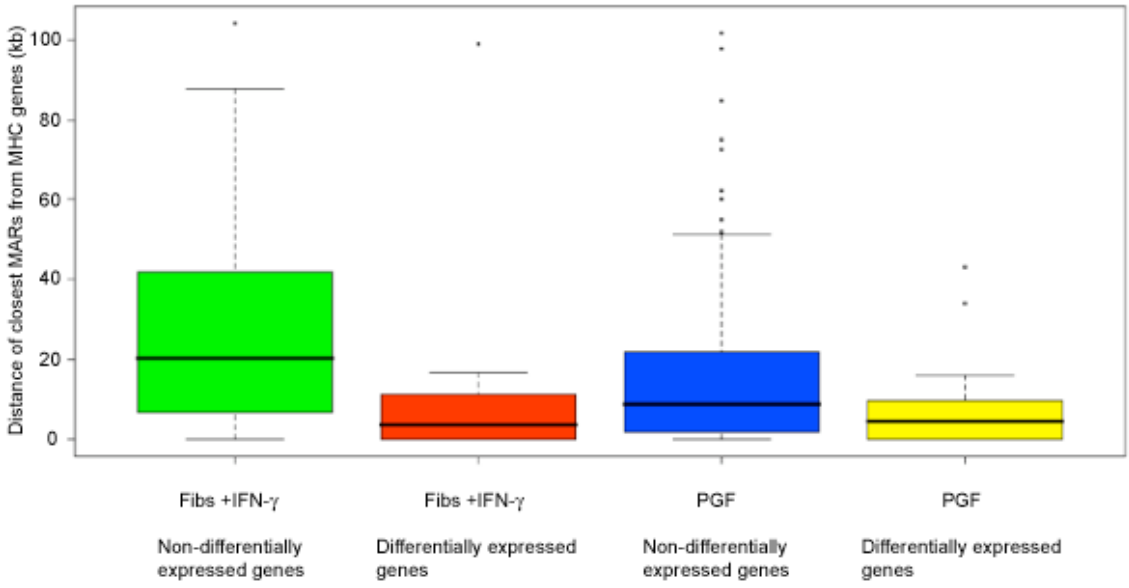

Figure 5.17 Distances of MARs to differentially and non-differentially expressed genes in IFN-γ MRC5 and PGF cells.

Boxplots show the interquartile range of the data, with the horizontal black lines running through each boxplot represent the median values. The whiskers extend as far as 1.5 times the interquartile range. Any values that lie beyond the whiskers are plotted as dots. A Mann-Whitney test showed that the differences between the distributions of differentially and non-differentially expressed genes were statistically significant for both cell types (p-values of 0.003 and 0.009, respectively).

5.2.8 Comparison between matrix attachment regions identified experimentally and *in silico* in IFN-γ induced MRC5 and PGF cells

The performance of three MAR prediction programmes SIDD, ChrClass and MARScan was assessed as described for untreated MRC5 cells.

The number of MARs predicted by all programmes is higher than that of experimentally identified MARs in IFN-γ induced MRC5 and PGF cells. It is important to note that SIDD, ChrClass and MARScan give extremely different distributions of MARs across the MHC locus. This is to be expected since each of these software programmes recognises different sequence motifs associated with MARs (described in Chapter 4).

Tables 5.3 and 5.4 show the number of predictions of experimentally defined MARs and not matching experimentally defined MARs in the two cell types. As seen in untreated MRC5 cells the number of false positives is significantly high, in both IFN-γ induced MRC5 and PGF cells. The number of true positives for ChrClass was approximately more than 2.5 times greater than SIDD and 3.5 times higher than MARScan. The number of false positive for ChrClass was slightly lower than SIDD and about 2.5 times higher than MARScan.

In agreement with the results from uninduced MRC5 cells, ChrClass outperformed SIDD and MARScan in both IFN-γ induced MRC5 and PGF cells. ChrClass had the highest sensitivity and specificity while SIDD and MARScan had the lowest specificity and sensitivity respectively.

In IFN-γ induced MRC5 cells, 15 MARs detected with the MHC microarrays were not predicted by any of the selected programs while in PGF that number was 17. As seen in the previous chapter, this suggests that additional sequence motifs associated with MARs should be incorporated into novel software programmes to improve their predicting power. These sequences could include specific classes of repetitive sequences such as LINEs since they are over-represented among MARs of the MHC relative to their abundance in non-MARs.

	ChrClass	SIDD	MARScan
Total Predicted	617	543	187
True Positive	191	75	55
False Positive	426	468	132
False Negative	18	71	69
Specificity	0.31	0.14	0.29
Sensitivity	0.86	0.34	0.25

Table 5.3 Comparison between different MAR predicting algorithms in IFN-γ treated cells.

	ChrClass	SIDD	MARScan
Total Predicted	617	543	187
True Positive	236	77	47
False Positive	381	476	140
False Negative	41	33	90
Specificity	0.38	0.14	0.25
Sensitivity	0.90	0.30	0.18

Table 5.4 Comparison between different MAR predicting algorithms in PGF cells.

5.3 Discussion

In order to examine the relationship between chromatin folding and gene expression, MARs were mapped in cells exhibiting differential expression of the MHC. By comparing the position of MARs in uninduced fibroblasts, IFN-γ induced fibroblasts and B lymphoblastoid cells, activation of the MHC was found to be accompanied by dynamic interactions between chromatin and the nuclear matrix. The vast recruitment of MARs to the nuclear matrix upon up-regulation of the MHC suggests that MARs play a role in transcriptional regulation of MHC genes.

As determined in uninduced fibroblasts, MARs in IFN-γ induced fibroblasts and B lymphoblastoid cells, MARs were found to be rich in AT and repetitive sequences, furthermore they correlate negatively with gene density. Despite the overall differences in MHC expression between uninduced, IFN-γ induced and B lymphoblastoid cells, MARs were mainly positioned within the MHC class I and class II regions. In all cell types examined, MARs predominantly occupy intergenic regions closely flanking several class I and class II genes and near MHC boundaries suggesting that MARs have an insulator function. This proposed role for MARs is further supported by recent findings showing that MARs shield transgenes from the effects of the neighbouring host chromatin domains (Girod et al., 2007; Halweg et al., 2005). It remains to be established whether MARs act as boundary elements by topologically constraining the DNA into loops or by forming a physical barrier that blocks interactions between promoter and unrelated enhancers.

Upon transcriptional up-regulation of MHC genes with IFN-γ, loop DNA adjacent to genes is recruited to the nuclear matrix. The number of MARs in IFN-γ induced fibroblasts is approximately two times greater than untreated fibroblasts and their average position is two times closer to differentially expressed MHC genes. The recruited MARs, mostly situated intergenically within the classical class I and II regions, reduce the average loop size across the MHC locus of IFN-γ induced fibroblasts. A similar pattern of anchoring and average loop size is seen in B lymphoblastoid cells, which constitutively express MHC genes that are up-regulated in IFN-γ induced fibroblasts. This association between transcriptional activation and recruitment of MARs has been previously reported at the murine Th2 cytokine locus

containing the co-ordinately regulated *Il4*, *Il3* and *Il5* genes within a gene cluster of 120 kb (Cai et al., 2006).

To gain a better understanding of MAR function in the MHC, the MARs of the different cell types were classified into three groups: structural, cell type specific and transcription-dependent (Heng et al., 2004; Heng et al., 2001; Shaposhnikov et al., 2007). Structural MARs are constitutively bound to the nuclear matrix in all cell types regardless of the transcriptional status of the MHC locus (Heng et al., 2004; Heng et al., 2001). MARs that bind to the nuclear matrix in all cell types could be critical in maintaining the higher-order chromatin structure of the MHC. Cell type specific MARs could reflect interactions between chromatin and cell type specific nuclear matrix proteins (Jackson et al., 1992). Studies on SATB1 have revealed that the binding of cell type specific nuclear matrix proteins at MARs is essential for tissue-specific regulation of gene expression (Cai et al., 2006; Galande et al., 2007). Therefore this group of MARs could interact with cell type specific matrix proteins, which are important for the distinct MHC expression profiles of fibroblasts and B lymphoblastoid cells. Transcription-dependent MARs are recruited to or released from the nuclear matrix based on the transcriptional status of a given genomic region (Heng et al., 2004). Most MARs of this group are recruited to the nuclear matrix upon activation of the MHC and may associate with a sub-class of MAR-binding proteins functioning as transcriptional activators of MHC genes.

These permanent and transient chromatin-matrix interactions should be examined in the context of previous findings showing that constitutively high gene expression or the induction of expression of certain genomic domains are accompanied by a higher-order re-structuring of chromatin (Brown et al., 2006; Chambeyron and Bickmore, 2004; Mahy et al., 2002; Volpi et al., 2000; Williams et al., 2002). In IFN-γ treated fibroblasts, the MHC emanates out of its CT forming giant chromatin loops with an increased frequency compared to untreated fibroblasts (Volpi et al., 2000). Here, FISH analysis revealed that the number of giant chromatin loops in B lymphoblastoid cells is largely equivalent to that seen in IFN-γ treated fibroblasts. Recent work by Christova et al. (2007) has shown that the formation of these giant loops is accompanied by chromatin decondensation, which probably provides a more ‘open’ chromatin environment (Christova et al., 2007). Decondensed chromatin could thus facilitate chromatin-matrix interactions by increasing accessibility for nuclear matrix proteins and transcription

machinery (Hart and Laemmli, 1998; Kimura et al., 1999). Since several components of transcription factories associate with the nuclear matrix, it is tempting to speculate that dynamic matrix-MAR interactions localise genes close to transcription factories optimising their regulation (Ottaviani et al., 2008b). Furthermore, giant chromatin loops may place the MHC in distinct nuclear microenvironments containing specific matrix proteins as well as transcription factors required for the expression of MHC genes. Taken together these findings indicate that there is an intimate relationship between gene expression, higher-order chromatin structure and the association of MARs with the nuclear matrix (Fig 5.18).

The role of MARs upon MHC induction could be explored in the future by interfering with their function and analysing the impact on regulation of gene expression. However, this approach cannot be used at present due to the lack of selective MAR-targeting agents to disrupt the binding of MARs with the nuclear matrix. A different strategy to analyse the structural and functional role of MARs includes the isolation and characterisation of MAR-binding proteins (Donev et al., 2002). If unknown, the function of these proteins could be analysed using approaches such as gene silencing. A similar strategy has been successfully applied to characterise the matrix protein SATB1 and the function of its MAR target sequences (Cai et al., 2003; Cai et al., 2006). Additional studies would also be required to understand whether MAR recruitment is an important early step in transcriptional activation of the MHC. This question can be addressed by time-course experiments to determine the kinetics of MAR recruitment to the nuclear matrix during IFN-γ activation.

The feasibility of large-scale MAR prediction was evaluated by comparing the positions of experimentally identified MARs in IFN-γ induced fibroblasts and B lymphoblastoid cells with those predicted by three MAR-prediction software programmes: ChrClass, SIDD and MARScan. As observed for uninduced fibroblasts in the previous chapter, ChrClass was found to have the best predicting power as measured by sensitivity and specificity. All the programmes used predicted a high number of false positives and therefore cannot yet substitute experimental verification. These findings show, however, that ChrClass could be a useful start where no experimental data is available. As seen, the binding of MARs to the nuclear matrix can be constitutive and transient, cell type specific and transcriptional dependent. The description of the sequence of these different MAR groups at the base pair level and the characterisation of target sites of

MAR-binding proteins will be essential for the development of improved MAR-predicting software.

The comparison between the position of MARs with CIITA binding sites and acetylated regions described in Chapter 3 did not reveal significant correlations. However, it is possible that such correlations were not detected since different cell types were used in these experiments.

In summary these results have shown that activation of MHC genes is accompanied by a massive re-organisation of chromatin architecture. The mapping of MARs in IFN-γ induced fibroblasts and B lymphoblastoid cells, and the comparison with those identified in uninduced fibroblasts has indicated that MARs are recruited to the nuclear matrix upon transcriptional up-regulation. MARs are formed adjacent to differentially expressed genes when the genes are activated suggesting that there is an intimate relationship between higher-order genome architecture and regulation of gene expression. This work thus provides a substantial framework for subsequent studies on DNA-protein interactions that are responsible for both the structural integrity of chromosomes and genomic function.

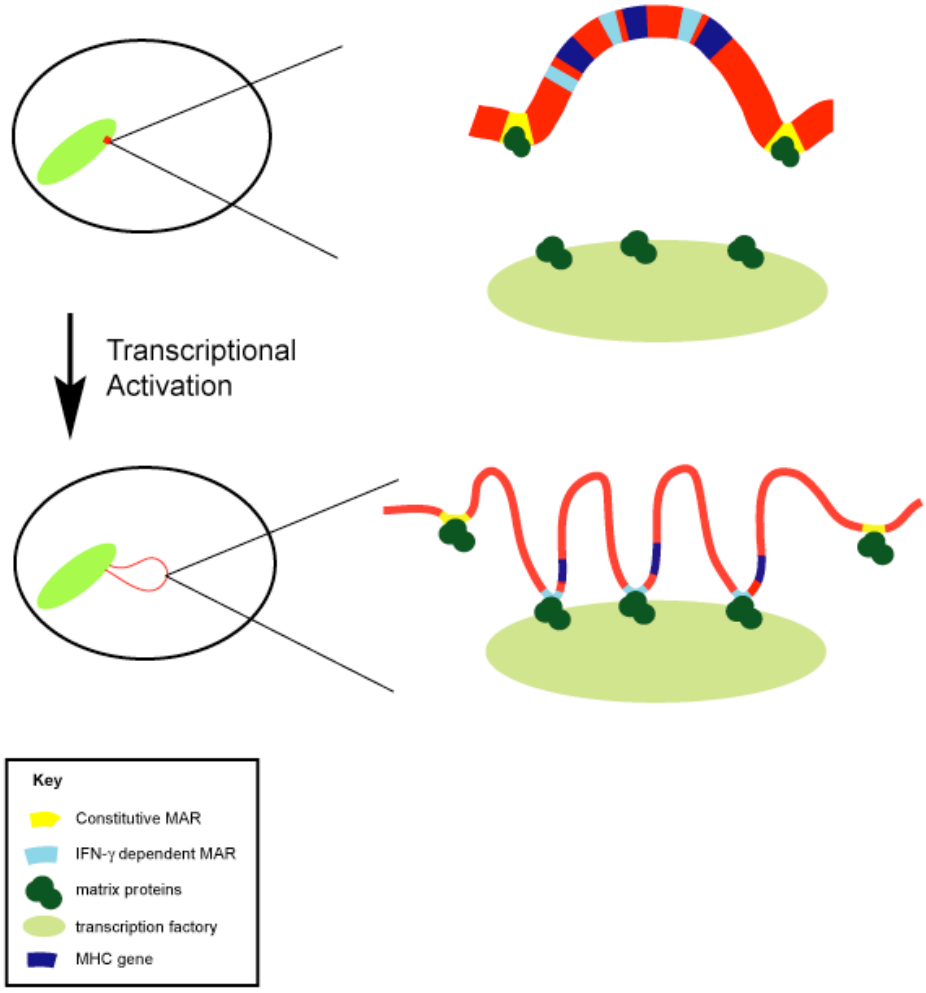

Figure 5.18 Model showing dynamic changes in chromatin organisation of the MHC locus upon transcriptional up-regulation.

In cells that are not expressing MHC class II genes, such as uninduced fibroblasts, the 4Mb MHC locus (red) is positioned within the chromosome 6 territory anchored by constitutive MARs (yellow) to nuclear matrix proteins (dark green), forming stable chromatin loops. Upon up-regulation of gene expression the entire MHC locus adopts a more 'open', decondensed chromatin structure, represented by the thinning and lengthening of the chromatin line, and the entire MHC region is repositioned outside of the CT. IFN-γ-dependent MARs (light blue) are recruited to the nuclear matrix, folding chromatin into smaller loops, and facilitating the interaction of chromatin and transcription factors (light green). Many IFN-γ-dependent MARs are positioned immediately adjacent to genes. The constitutive MARs remain attached to nuclear matrix proteins following transcriptional up-regulation and some may have an insulatory function.

6 Summary and Discussion

In this thesis, I have examined the relationship between genomic anchoring and gene expression in the MHC, a large human locus containing clusters of co-ordinately regulated genes. Using a novel tiling path array of the MHC, I have shown that MARs are unevenly distributed across the MHC region. They are mainly situated within intergenic domains and closely flank several genes, including those involved in antigen processing and presentation, as well as the boundaries defining the MHC sub-regions. Although a unique consensus was not identified, MARs have a high AT and repeat sequence content. Analysis of cell types with different expression profiles suggests that transcriptional up-regulation of the MHC is accompanied by a significant recruitment of MARs to the nuclear matrix. Of particular interest, certain MARs are formed close to transcriptionally up-regulated genes upon MHC activation, while others are permanently associated with the nuclear matrix in all cell types. Collectively, these findings suggest that the rearrangement of chromatin architecture to form new loop domains may be an integral component of gene activation.

6.1 Higher-order chromatin structure

What does the map of MARs across the MHC in different cell types tell us about chromatin architecture, and how compatible is it with the models described in Chapter 1? To address these questions, one should consider the spatial arrangement of MARs in relation to the proposed chromatin loop domain. The common view is that MARs are located at the bases of the chromatin loop structures. However, each MAR may not identify the base of a separate loop, since consecutive, closely juxtaposed MARs may together be positioned at the base of a loop by their common attachment to a particular set of nuclear matrix proteins. Certain MARs may be located along a loop rather than exclusively at the base, and furthermore, one cannot exclude a role for MARs in anchoring chromatin folds, rather than loops.

With respect to the MHC, MARs identified within a 300 kb region of the class I region have been shown by a chromatin ligation assay, to be present at the bases of loops, with each pair of MARs that form a loop structure being in close spatial proximity *in vivo*

(Kumar et al., 2007a). In contrast, 3C analysis of the *Ifng* locus indicates that certain MARs are associated with a linear, rather than looped, chromatin conformation, indicating that not all MARs mediate the organisation of chromatin into loop structures (Eivazova et al., 2007). The methodology used in this thesis to isolate the nuclear matrix is identical to that shown to extract proteins with both structural functions, such as lamins, and no known structural functions, such as general transcription factors (Noma et al., 2006). Certain MARs in the MHC might thus only have a regulatory function and be located away from the bases of the loop structures. Analysis of the three-dimensional chromatin architecture using 3C analysis needs to be performed to determine which MARs mediate loop formation and which have a regulatory function.

The findings from the present study should be examined in light of the different models of higher-order chromatin structure. The MLS model postulates that successive bases of chromatin loops are held together by 'loop base springs' (Munkel et al., 1999). However, as the existence of a proteinaceous nuclear scaffold tethering the chromatin is not excluded, the MARs of the MHC could mediate the organisation of the MHC into rosette-like structures containing chromatin loops. If all the MARs are located at the bases of the loop structures, the MHC class I and II regions would be organised into smaller chromatin loops (~15 kb) compared to those of the MHC class III and the extended class II regions (~80 kb). In contrast, if certain MARs of the MHC class I and II regions were located away from the bases of the loop structures, the sizes of the chromatin loops could be similar across the MHC sub-regions.

The RW/GL model predicts the formation of chromatin loops in the order of several Mb attached to a proteinaceous nuclear scaffold. According to this model, the MHC should only contain very few MARs and not the large number that were identified experimentally in all three cell types. Although this model might not be suitable to interpret our results, smaller chromatin loops might still be formed within the relatively large loops and allow for the existence of a higher number of MARs (Sachs et al., 1995).

The chromonema fibre model does not appear to agree with our findings since it favours the formation of chromatin folds maintained by chromatin interactions. However, this model can accommodate a nuclear matrix mediating the tethering of the chromatin folds (Belmont and Bruce, 1994). In this instance, MARs would organise the MHC class I

and II regions into more tightly folded chromatin compared to the MHC class III and extended class II regions.

6.2 Relationship to giant chromatin loops

Within the nucleus, chromatin exists in open and closed configurations. Certain genomic regions are subject to dynamic changes in conformation, according to their transcriptional status and other parameters. None of the models described above provide information on how the different levels of chromatin condensation are established or how they are altered upon interaction with regulatory nuclear machineries.

The formation of giant chromatin loops carrying the MHC upon transcriptional activation has been described earlier (Volpi et al., 2000). It might be expected that the release of a chromatin loop away from the chromosome territory would be achieved through release of MARs upon transcriptional activation. Rather, what is seen is the maintenance of virtually all existing MARs, together with the formation of additional MARs indicating new sites of anchoring to the nuclear matrix. This observation could reflect the biological and physical properties of the nuclear matrix. If nuclear matrix components move within the nucleus they might facilitate the formation of giant chromatin loops through their interaction with MARs. This is in agreement with well-established findings that many proteins in the nucleus are highly mobile and generate a dynamic, architectural framework, where nuclear processes take place. Immediately after treatment of cells with IFN-γ and activation of the JAK-STAT pathway, phosphorylated STAT1 moves rapidly into the nucleus. As the giant chromatin loops appear within minutes, and only form if STAT1 is present (Christova et al., 2007), one could speculate that STAT1 may be involved in MAR-matrix interactions that generate large-scale reconfiguration of chromatin architecture.

MARs are likely to mediate the tethering of the local chromatin loops across the MHC, however, whether these genomic sequences are also involved in the formation of the giant chromatin loops remains to be established. The average size of the giant chromatin loops is in the order of several Mb (Volpi et al., 2000), while the average distance between MARs in the MHC is approximately several kb. The formation of additional MARs upon activation of the MHC could indicate the formation of new chromatin

loops, as observed in the mouse Th2 cytokine locus (Cai et al., 2006). However, this is difficult to reconcile with the observed decondensation of the MHC that accompanies its transcriptional up-regulation (Christova et al., 2007). It is more likely that the overall decondensation of the locus increases access of MARs, which are located a distance away from loop bases, to nuclear matrix proteins. Thus, the associations of these MARs with the nuclear matrix would not result in the formation of additional chromatin loops. Another possibility is that different MARs might be responsible for different levels of chromatin compaction so that the release of a few selected MARs might still lead to chromatin decondensation despite an overall recruitment of MARs. The data are thus consistent, firstly, with a complex hierarchy of DNA folding to achieve effective packaging of the MHC, and secondly, with activation of MHC genes being associated with conformational changes occurring at multiple levels.

6.3 Insights into the nature of the nuclear matrix

As this thesis concerns the investigation of MARs, it is appropriate to evaluate whether the nuclear matrix is a true nuclear entity and analyse what the data reveal about the composition and properties of the nuclear matrix.

The extraction of soluble histone and non-histone proteins with high-salt buffers can form artifactual protein complexes since the negatively charged polyanions of the buffers interact with the cationic sites of proteins (Pederson, 2000). This process may disrupt both pre-existing DNA-protein interactions and protein-protein interactions and lead to the formation of new aggregates. This limitation of the methodology has to be taken into account when considering recent findings proteins SATB1, topoisomerase II and condensin, which are found in the matrix fraction of the nucleus following treatment with 2 M NaCl buffer (Maeshima and Laemmli, 2003; Seo et al., 2005). All three proteins are important in chromatin architecture. SATB1 has a ‘cage-like’ network distribution pattern in interphase nuclei (Cai et al., 2003; Cai et al., 2006). Topoisomerase II and condensin form two juxtaposed or coiled chain structures along the longitudinal axis of metaphase chromosomes (Kireeva et al., 2004; Maeshima and Laemmli, 2003). These findings strongly suggest that components of the nuclear matrix form complex structures in nuclei *in vivo.* However, additional components might simply co-precipitate in the matrix fraction as a result of the high-salt treatment and thus

might not actually participate in the maintenance of chromosomal structure. In order to distinguish the proteins of the nuclear matrix from the artifactual ones, more components need to be characterised and studied using techniques such as ChIP and *in vivo* imaging.

The presence of both constitutive and dynamic MARs in the MHC is likely to reflect the heterogeneous composition of the nuclear matrix, which is constituted not only of structural proteins but also of factors involved in transcription. The nuclear matrix contains a few structural components including SATB1 and lamins, which can self-assemble to form complex higher-order nuclear structures extending throughout the nuclear interior (Cai et al., 2006; Dechat et al., 2008). Most nuclear matrix components, including chromatin remodelling enzymes and general transcription factors, have a diffuse distribution in the nucleus of living cells (Bettinger et al., 2004; Noma et al., 2006; Reyes et al., 1997). The nuclear matrix thus does not appear to be a continuous structural framework but rather is a collection of different proteins that mediate high-order chromosome structure and regulate gene expression in response to transcriptional up-regulation.

6.4 The architecture of the nucleus

An understanding of chromatin architecture has to incorporate the surrounding nuclear environment. The Inter-Chromatin Domain (ICD) model describes the nucleus as having compartmentalised chromatin domains separated by a network of chromatin-free space (Cremer and Cremer, 2001). This view of the nucleus is derived from microscopic imaging of the CTs and ICD. Experiments showing that approximately 90% of DNA, corresponding to about 50% of the nuclear mass, can be removed by extraction whilst preserving the structure of the ICD, suggest that the nuclear matrix could be situated within the channels of the ICD (Jackson, 2003). This hypothesis is further supported by the findings that both the ICD and the nuclear matrix are implicated in transcription and replication and are associated with nuclear components such as PML and Cajal bodies, which have a role in mRNA maturation and RNP assembly (Cremer et al., 2004; Gribbon et al., 2002; Zhong et al., 2000).

The definition of the ICD as the chromatin-free compartment separating CTs and chromatin domains is still controversial. Intermingling between CTs indicates that the ICD might not be preferentially distributed at the borders of CTs (Branco and Pombo, 2006). In addition, experiments showing that large macromolecules are able to diffuse within condensed chromatin suggest that the ICD might ramify in between fine chromatin domains (Cremer and Cremer, 2001; Verschure et al., 2003). These data indicate that rather than simply being compartmentalised into chromatin and chromatin-free domains, the nucleus is divided into different compartments characterised by variable volumes of chromatin-free space. Since it is highly condensed, heterochromatin contains less chromatin-free space than euchromatin, however, chromatin-free space is found throughout both domains. This network of chromatin-free space might be filled with different molecules including components of the nuclear matrix proteins, which are able to organise chromosomal architecture through their interactions with MARs.

In order to understand the different elements responsible for the compartmentalisation of the nucleus it is also necessary to explore the forces generated by molecular crowding, which are strong in the nucleus due to the high macromolecule concentration (in the range of 100 mg/ml) (Hancock, 2004b). These forces vastly increase the association constants of intermolecular interactions and might play a role in the assembly and function of distinct nuclear compartments including PML bodies and nucleoli (Hancock, 2004b). There is experimental evidence showing that these compartments disassembled with dispersion of their molecular components when nuclei expand in volume in hypotonic buffer and reassembled when the nuclear volume returns to normal in physiological buffers (Hancock, 2004a).

6.5 Genomic features of matrix attachment regions

The average AT content of MARs in all cell types analysed here is equal to or higher than 58% while that of non-MAR DNA is equal to or lower than 53%. MARs are therefore relatively AT-rich, confirming the results of previous small-scale investigations (Purbowasito et al., 2004; Shaposhnikov et al., 2007). These findings could have important practical implications since they suggest that drugs such as bizelesin, which bind selectively to AT-rich regions, could exhibit their cytotoxicity by interfering with the function of MARs (as discussed in Chapter 1). However, MARs

with low AT content have also been found indicating that high AT content alone is not sufficient for interaction of a genomic sequence with the nuclear matrix.

MARs are relatively rich in repetitive elements since the average content of repetitive sequences of MARs is equal to or higher than 49% while that of non-MAR DNA is equal to or lower than 43%. A positive correlation between MARs and repetitive elements was indicated previously in studies of several genomic regions including the human serpin gene cluster and a 1 Mb imprinted domain on mouse chromosome 7 (Purbowasito et al., 2004; Rollini et al., 1999). Further sequence analysis revealed that LINEs are the predominant class of repetitive sequences in the MARs in the MHC, and are consistently over-represented relative to their overall abundance in non-MAR DNA. This indicates that LINEs might contribute to the binding of DNA to the nuclear matrix and thus play a significant role in chromatin architecture.

In all cell types studied here, intragenic sequences constitute approximately 40% of non-MAR DNA and 20% for MARs. In addition, MARs show a negative correlation with gene density. The preferential localisation of MARs within intergenic sequences as well as their proximity to MHC genes and class boundaries suggest that they might function as boundary elements. This is also supported by previous transgene experiments revealing that MARs are able to shield genes from positioning effects (Girod et al., 2007; Halweg et al., 2005). Although the details need to be established, this insulatory function might arise from the ability of MARs of forming topologically constrained looped structures which protect the intervening DNA from the influence of chromatin outside the loop.

Another feature of MARs observed consistently in all three cell types is that they are approximately two times less conserved than non-MAR DNA. No unique sequence consensus was identified for MARs, suggesting that their conserved function across different species is not dictated by an evolutionary conserved primary sequence but rather by diverse DNA motifs generating distinct higher-order chromatin structures. A similar finding was shown for methylated sites on histones, which were found to be functionally conserved between the human and mouse genomes at orthologous sites even when their underlying sequence was not highly conserved (Bernstein et al., 2005).

The absence of a unique consensus characterising MARs has led to difficulties in predicting the position of MARs using *in silico* approaches. The performance of the computer programmes SIDD, ChrClass and MARScan was assessed in this thesis using experimental data from the three cell types. These programmes are based on different sets of DNA motifs associated with MARs, thus, it is not surprising that their predictions across the MHC overlap only partially. Although the performance of each programme was found to vary in the three cell types, Chrclass outperformed SIDD and MARScan since it had the highest sensitivity and specificity. In each cell type, a significant number of MARs were detected experimentally that were not predicted by any of the selected programmes. The positive and negative correlations between MARs and the genomic features analysed, which include AT content, repetitive sequences and gene density, should be incorporated into these programmes to improve their predicting power.

6.6 The structural and functional relationship between matrix attachment regions and the MHC

The efficient up-regulation of expression of MHC genes upon specific stimulatory signals regulates antigen presentation to T cells and ultimately determines the effectiveness of the immune response. The recruitment of genomic anchors near differentially expressed MHC genes accompanies their transcriptional activation, indicating that they could be critical for the regulation of the MHC locus.

The MHC class I and II regions contain a larger number of MARs and are therefore more topologically constrained than the class III and the extended class II regions. Since the class I and II regions have significantly higher AT and repetitive element content as well as higher gene density than the class III and extended class II regions, the uneven distribution of MARs might be due to the distinct genomic features of the MHC classes. The similar MAR distribution between the class I and class II regions might also reflect their similar evolution. The organisation and sequence of the genes contained in these classes suggest that they are derived from a common ancestor (Kasahara et al., 1995; Klein and Sato, 1998). In contrast, most of the class III genes and extended class II genes are not functionally or evolutionarily related to one another (Sultmann et al., 2000).

The organisation of chromatin into repeating local loops tethered to the nuclear scaffold might define distinct chromatin domains and constitute the basic structural unit of chromatin for regulating DNA transcription and replication (Bode et al., 1995; Heng et al., 2001; Nickerson et al., 1995). Furthermore, if the MHC sub-regions are organised in a rosette-like conformation as proposed by the MLS model, several MARs could be tethered at the centre of each rosette and multiple loops could share the same nuclear microenvironment facilitating the co-ordinate regulation of the MHC genes contained within the class I and II regions (Munkel et al., 1999).

One of the most striking features of the human genome is its subdivision into G- and R-bands (as discussed in Chapter 1). In the MHC, the class II region replicates late and it is highly AT-rich (Tenzen et al., 1997). We would therefore expect the MHC class II region to correspond to a G-band. Does the MAR distribution provide insights into the banding phenomenon? According to Saitoh and Laemmli (1994), G- and R-bands arise from the differential compaction of MARs along metaphase chromosomes (Saitoh and Laemmli, 1994). They showed that G bands correspond to regions where MARs are more tightly coiled than R-bands. A similar pattern might also exist in interphase chromosomes as our findings suggest that the MHC class II contains a significant higher number of MARs than the adjacent sub-regions and might thus be subjected to higher structural constraints.

Previous small-scale studies on different organisms have indicated that dynamic recruitment of MARs to the nuclear matrix regulates the expression of genes involved in cellular differentiation and disease development as well as additional nuclear processes including DNA replication and mitosis (discussed in Chapter 1). The expression of several MHC proteins, including HLA class I molecules, is low in pluripotent human embryonic stem cells but shows a significant increase during differentiation (Drukker and Benvenisty, 2004; Drukker et al., 2002). In order to improve our understanding of the molecular mechanisms regulating MHC gene expression during cellular differentiation, the MHC tiling path arrays could be used to map the MARs in cells at different developmental stages. The identification of MARs in cells exhibiting abnormal MHC gene expression could determine whether MARs contribute to the diseases associated with this locus. In preparation for mitosis, chromatin starts undergoing profound conformational changes to reach the extreme

degree of compaction observed in metaphase chromosomes. The isolation of MARs from metaphase nuclei and their localisation across the MHC could reveal whether the rearrangement of MARs is also involved in the formation of metaphase chromosomes. In addition to consolidating the findings of the present study, these experiments will provide a broader view of the relationships between structure and function that underlie chromosome biology.

6.7 Final Remarks

Our knowledge of how the MHC functions in the context of the nucleus has been enhanced by the findings that distinct genomic sites of the MHC interact with proteins of the nuclear matrix. These MARs are likely to regulate the expression of the MHC locus by dynamically organising chromatin in three-dimensional space. The identification of CIITA-binding sites and histone acetylated regions across the MHC emphasises that this critical region of the genome is regulated at multiple levels. The characterisation of *cis*-regulatory elements, regulatory proteins and epigenetic modifications as well as their co-ordination and compartmentalisation into distinct nuclear microenvironments will allow us to gain a more comprehensive view of genomic function and identify truly novel targets for therapy.

References

Adachi, Y., E. Kas, and U.K. Laemmli. 1989. Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. *Embo J*. 8:3997-4006.

Agalioti, T., G. Chen, and D. Thanos. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell*. 111:381-92.

Aggarwal, B.D., and B.R. Calvi. 2004. Chromatin regulates origin activity in Drosophila follicle cells. *Nature*. 430:372-6.

Ahn, S.H., R.L. Diaz, M. Grunstein, and C.D. Allis. 2006. Histone H2B deacetylation at lysine 11 is required for yeast apoptosis induced by phosphorylation of H2B at serine 10. *Mol Cell*. 24:211-20.

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. 1994. Molecular biology of the cell. Garland Publishing, New York.

Alley, M.C., M.G. Hollingshead, C.M. Pacula-Cox, W.R. Waud, J.A. Hartley, P.W. Howard, S.J. Gregson, D.E. Thurston, and E.A. Sausville. 2004. SJG-136 (NSC 694501), a novel rationally designed DNA minor groove interstrand cross-linking agent with potent and broad spectrum antitumor activity: part 2: efficacy evaluations. *Cancer Res*. 64:6700-6.

Alvarado-Guerri, R., C.M. Cabrera, F. Garrido, and M.A. Lopez-Nevot. 2005. TAP1 and TAP2 polymorphisms and their linkage disequilibrium with HLA-DR, -DP, and -DQ in an eastern Andalusian population. *Hum Immunol*. 66:921-30.

Alvarez, J.D., D.H. Yasui, H. Niida, T. Joh, D.Y. Loh, and T. Kohwi-Shigematsu. 2000. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev*. 14:521-35.

Anachkova, B., V. Djeliova, and G. Russev. 2005. Nuclear matrix support of DNA replication. *J Cell Biochem*. 96:951-61.

Ansari, A., and M. Hampsey. 2005. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*. 19:2969-78.

Apostolou, E., and D. Thanos. 2008. Linking differential chromatin loops to transcriptional decisions. *Mol Cell*. 29:154-6.

Aquino-Galvez, A., A. Camarena, M. Montano, A. Juarez, A.C. Zamora, G. Gonzalez-Avila, M. Checa, G. Sandoval-Lopez, G. Vargas-Alarcon, J. Granados, A. Pardo, J. Zuniga, and M. Selman. 2008. Transporter associated with antigen processing (TAP) 1 gene polymorphisms in patients with hypersensitivity pneumonitis. *Exp Mol Pathol*. 84:173-7.

Baker, S.P., and P.A. Grant. 2007. The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene*. 26:5329-40.

Bakke, O., and B. Dobberstein. 1990. MHC class II-associated invariant chain contains a sorting signal for endosomal compartments. *Cell*. 63:707-16.

Bednar, J., R.A. Horowitz, S.A. Grigoryev, L.M. Carruthers, J.C. Hansen, A.J. Koster, and C.L. Woodcock. 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*. 95:14173-8.

Belmont, A.S., and K. Bruce. 1994. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol*. 127:287-302.

Benham, C., T. Kohwi-Shigematsu, and J. Bode. 1997. Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. *J Mol Biol*. 274:181-96.

Bentley, D.L. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol*. 17:251-6.

Beresford, G.W., and J.M. Boss. 2001. CIITA coordinates multiple histone acetylation modifications at the HLA-DRA promoter. *Nat Immunol*. 2:652-7.

Berezney, R. 2002. Regulating the mammalian genome: the role of nuclear architecture. *Adv Enzyme Regul*. 42:39-52.
Berezney, R., and D.S. Coffey. 1974. Identification of a nuclear protein matrix. *Biochem Biophys Res Commun*. 60:1410-7.
Berezney, R., and D.S. Coffey. 1977. Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei. *J Cell Biol*. 73:616-37.
Berezney, R., M.J. Mortillaro, H. Ma, X. Wei, and J. Samarabandu. 1995. The nuclear matrix: a structural milieu for genomic function. *Int Rev Cytol*. 162A:1-65.
Berger, F., and V. Gaudin. 2003. Chromatin dynamics and Arabidopsis development. *Chromosome Res*. 11:277-304.
Bernstein, B.E., E.L. Humphrey, R.L. Erlich, R. Schneider, P. Bouman, J.S. Liu, T. Kouzarides, and S.L. Schreiber. 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A*. 99:8695-700.
Bernstein, B.E., M. Kamal, K. Lindblad-Toh, S. Bekiranov, D.K. Bailey, D.J. Huebert, S. McMahon, E.K. Karlsson, E.J. Kulbokas, 3rd, T.R. Gingeras, S.L. Schreiber, and E.S. Lander. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 120:169-81.
Berrios, M., N. Osheroff, and P.A. Fisher. 1985. In situ localization of DNA topoisomerase II, a major polypeptide component of the Drosophila nuclear matrix fraction. *Proc Natl Acad Sci U S A*. 82:4142-6.
Bettinger, B.T., D.M. Gilbert, and D.C. Amberg. 2004. Actin up in the nucleus. *Nat Rev Mol Cell Biol*. 5:410-5.
Bickmore, W.A., and K. Oghene. 1996. Visualizing the spatial relationships between defined DNA sequences and the axial region of extracted metaphase chromosomes. *Cell*. 84:95-104.
Bjorkman, P.J., and P. Parham. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem*. 59:253-88.
Bjorkman, P.J., M.A. Saper, B. Samraoui, W.S. Bennett, J.L. Strominger, and D.C. Wiley. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*. 329:506-12.
Blanton, J., M. Gaszner, and P. Schedl. 2003. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev*. 17:664-75.
Blencowe, B.J., J.A. Nickerson, R. Issner, S. Penman, and P.A. Sharp. 1994. Association of nuclear matrix antigens with exon-containing splicing complexes. *J Cell Biol*. 127:593-607.
Bode, J., S. Goetze, H. Heng, S.A. Krawetz, and C. Benham. 2003. From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res*. 11:435-45.
Bode, J., Y. Kohwi, L. Dickinson, T. Joh, D. Klehr, C. Mielke, and T. Kohwi-Shigematsu. 1992. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*. 255:195-7.
Bode, J., T. Schlake, M. Rios-Ramirez, C. Mielke, M. Stengert, V. Kay, and D. Klehr-Wirth. 1995. Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int Rev Cytol*. 162A:389-454.
Boehm, U., T. Klamp, M. Groot, and J.C. Howard. 1997. Cellular responses to interferon-gamma. *Annu Rev Immunol*. 15:749-95.
Bolzer, A., G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Muller, R. Eils, C. Cremer, M.R. Speicher, and T. Cremer. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*. 3:e157.
Bonne, G., M.R. Di Barletta, S. Varnous, H.M. Becane, E.H. Hammouda, L. Merlini, F. Muntoni, C.R. Greenberg, F. Gary, J.A. Urtizberea, D. Duboc, M. Fardeau, D.

Toniolo, and K. Schwartz. 1999. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat Genet*. 21:285-8.

Botuyan, M.V., J. Lee, I.M. Ward, J.E. Kim, J.R. Thompson, J. Chen, and G. Mer. 2006. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell*. 127:1361-73.

Boulikas, T. 1993. Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J Cell Biochem*. 52:14-22.

Boulikas, T. 1995. Chromatin domains and prediction of MAR sequences. *Int Rev Cytol*. 162A:279-388.

Boyle, S., S. Gilchrist, J.M. Bridger, N.L. Mahy, J.A. Ellis, and W.A. Bickmore. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet*. 10:211-9.

Branco, M.R., and A. Pombo. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol*. 4:e138.

Branco, M.R., and A. Pombo. 2007. Chromosome organization: new facts, new models. *Trends Cell Biol*. 17:127-34.

Brown, D.T. 2001. Histone variants: are they functionally heterogeneous? *Genome Biol*. 2:REVIEWS0006.

Brown, J.M., J. Leach, J.E. Reittie, A. Atzberger, J. Lee-Prudhoe, W.G. Wood, D.R. Higgs, F.J. Iborra, and V.J. Buckle. 2006. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*. 172:177-87.

Byrd, K., and V.G. Corces. 2003. Visualization of chromatin domains created by the gypsy insulator of Drosophila. *J Cell Biol*. 162:565-74.

Cai, S., H.J. Han, and T. Kohwi-Shigematsu. 2003. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet*. 34:42-51.

Cai, S., and T. Kohwi-Shigematsu. 1999. Intranuclear relocalization of matrix binding sites during T cell activation detected by amplified fluorescence in situ hybridization. *Methods*. 19:394-402.

Cai, S., C.C. Lee, and T. Kohwi-Shigematsu. 2006. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*. 38:1278-88.

Capco, D.G., K.M. Wan, and S. Penman. 1982. The nuclear matrix: three-dimensional architecture and protein composition. *Cell*. 29:847-58.

Carrington, M., G.W. Nelson, M.P. Martin, T. Kissner, D. Vlahov, J.J. Goedert, R. Kaslow, S. Buchbinder, K. Hoots, and S.J. O'Brien. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science*. 283:1748-52.

Carter, C.A., W.R. Waud, L.H. Li, T.F. DeKoning, J.P. McGovren, and J. Plowman. 1996. Preclinical antitumor activity of bizelesin in mice. *Clin Cancer Res*. 2:1143-9.

Celic, I., H. Masumoto, W.P. Griffith, P. Meluh, R.J. Cotter, J.D. Boeke, and A. Verreault. 2006. The sirtuins hst3 and Hst4p preserve genome integrity by controlling histone h3 lysine 56 deacetylation. *Curr Biol*. 16:1280-9.

Chambeyron, S., and W.A. Bickmore. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*. 18:1119-30.

Chernov, I.P., S.B. Akopov, L.G. Nikolaev, and E.D. Sverdlov. 2002. Identification and mapping of nuclear matrix-attachment regions in a one megabase locus of human chromosome 19q13.12: long-range correlation of S/MARs and gene positions. *J Cell Biochem*. 84:590-600.

Cheung, P., C.D. Allis, and P. Sassone-Corsi. 2000. Signaling to chromatin through histone modifications. *Cell*. 103:263-71.
Choo, S.Y. 2007. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J*. 48:11-23.
Christova, R., T. Jones, P.J. Wu, A. Bolzer, A.P. Costa-Pereira, D. Watling, I.M. Kerr, and D. Sheer. 2007. P-STAT1 mediates higher-order chromatin remodelling of the human MHC in response to IFNgamma. *J Cell Sci*. 120:3262-70.
Christova, R., and T. Oelgeschlager. 2002. Association of human TFIID-promoter complexes with silenced mitotic chromatin in vivo. *Nat Cell Biol*. 4:79-82.
Chuang, C.H., A.E. Carpenter, B. Fuchsova, T. Johnson, P. de Lanerolle, and A.S. Belmont. 2006. Long-range directional movement of an interphase chromosome site. *Curr Biol*. 16:825-31.
Clayton, A.L., S. Rose, M.J. Barratt, and L.C. Mahadevan. 2000. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *Embo J*. 19:3714-26.
Collins, J.M., and A.K. Chu. 1987. Binding of the DNA polymerase alpha-DNA primase complex to the nuclear matrix in HeLa cells. *Biochemistry*. 26:5600-7.
Colonna, M., M. Bresnahan, S. Bahram, J.L. Strominger, and T. Spies. 1992. Allelic variants of the human putative peptide transporter involved in antigen processing. *Proc Natl Acad Sci U S A*. 89:3932-6.
Cook, P.R. 1999. The organization of replication and transcription. *Science*. 284:1790-5.
Cook, P.R. 2002. Predicting three-dimensional genome structure from transcriptional activity. *Nat Genet*. 32:347-52.
Cremer, M., J. von Hase, T. Volm, A. Brero, G. Kreth, J. Walter, C. Fischer, I. Solovei, C. Cremer, and T. Cremer. 2001. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res*. 9:541-67.
Cremer, T., and C. Cremer. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. 2:292-301.
Cremer, T., and C. Cremer. 2006. Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem*. 50:223-72.
Cremer, T., C. Cremer, H. Baumann, E.K. Luedtke, K. Sperling, V. Teuber, and C. Zorn. 1982. Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet*. 60:46-56.
Cremer, T., K. Kupper, S. Dietzel, and S. Fakan. 2004. Higher order chromatin architecture in the cell nucleus: on the way from structure to function. *Biol Cell*. 96:555-67.
Cresswell, P. 1996. Invariant chain structure and MHC class II function. *Cell*. 84:505-7.
Croft, J.A., J.M. Bridger, S. Boyle, P. Perry, P. Teague, and W.A. Bickmore. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*. 145:1119-31.
Cullen, K.E., M.P. Kladde, and M.A. Seyfred. 1993. Interaction between transcription regulatory regions of prolactin chromatin. *Science*. 261:203-6.
Darley, R., A. Morris, J. Passas, and W. Bateman. 1993. Interactions between interferon gamma and retinoic acid with transforming growth factor beta in the induction of immune recognition molecules. *Cancer Immunol Immunother*. 37:112-8.

Darnell, J.E., Jr., I.M. Kerr, and G.R. Stark. 1994. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*. 264:1415-21.
Das, A.T., M.E. Luderus, and W.H. Lamers. 1993. Identification and analysis of a matrix-attachment region 5' of the rat glutamate-dehydrogenase-encoding gene. *Eur J Biochem*. 215:777-85.
Dausset, J. 1958. [Iso-leuko-antibodies.]. *Acta Haematol*. 20:156-66.
David-Watine, B., A. Israel, and P. Kourilsky. 1990. The regulation and expression of MHC class I genes. *Immunol Today*. 11:286-92.
Davie, J.R., and M.J. Hendzel. 1994. Multiple functions of dynamic histone acetylation. *J Cell Biochem*. 55:98-105.
Deakin, J.E., A.T. Papenfuss, K. Belov, J.G. Cross, P. Coggill, S. Palmer, S. Sims, T.P. Speed, S. Beck, and J.A. Graves. 2006. Evolution and comparative analysis of the MHC Class III inflammatory region. *BMC Genomics*. 7:281.
Dechat, T., K. Pfleghaar, K. Sengupta, T. Shimi, D.K. Shumaker, L. Solimando, and R.D. Goldman. 2008. Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev*. 22:832-53.
Deininger, M.H., R. Meyermann, and H.J. Schluesener. 2002. The allograft inflammatory factor-1 family of proteins. *FEBS Lett*. 514:115-21.
Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing chromosome conformation. *Science*. 295:1306-11.
Dillon, N., and R. Festenstein. 2002. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet*. 18:252-8.
Dimitrova, D.S., I.T. Todorov, T. Melendy, and D.M. Gilbert. 1999. Mcm2, but not RPA, is a component of the mammalian early G1-phase prereplication complex. *J Cell Biol*. 146:709-22.
Djeliova, V., G. Russev, and B. Anachkova. 2001a. Distribution of DNA replication origins between matrix-attached and loop DNA in mammalian cells. *J Cell Biochem*. 80:353-9.
Djeliova, V., G. Russev, and B. Anachkova. 2001b. Dynamics of association of origins of DNA replication with the nuclear matrix during the cell cycle. *Nucleic Acids Res*. 29:3181-7.
Dobrzycka, K.M., K. Kang, S. Jiang, R. Meyer, P.H. Rao, A.V. Lee, and S. Oesterreich. 2006. Disruption of scaffold attachment factor B1 leads to TBX2 up-regulation, lack of p19ARF induction, lack of senescence, and cell immortalization. *Cancer Res*. 66:7859-63.
Donev, R.M. 2000. The type of DNA attachment sites recovered from nuclear matrix depends on isolation procedure used. *Mol Cell Biochem*. 214:103-10.
Donev, R.M., T.A. Doneva, W.R. Bowen, and D. Sheer. 2002. HnRNP-A1 binds directly to double-stranded DNA in vitro within a 36 bp sequence. *Mol Cell Biochem*. 233:181-5.
Drukker, M., and N. Benvenisty. 2004. The immunogenicity of human embryonic stem-derived cells. *Trends Biotechnol*. 22:136-41.
Drukker, M., G. Katz, A. Urbach, M. Schuldiner, G. Markel, J. Itskovitz-Eldor, B. Reubinoff, O. Mandelboim, and N. Benvenisty. 2002. Characterization of the expression of MHC proteins in human embryonic stem cells. *Proc Natl Acad Sci U S A*. 99:9864-9.
Dunn, K.L., H. Zhao, and J.R. Davie. 2003. The insulator binding protein CTCF associates with the nuclear matrix. *Exp Cell Res*. 288:218-23.
Earnshaw, W.C., B. Halligan, C.A. Cooke, M.M. Heck, and L.F. Liu. 1985. Topoisomerase II is a structural component of mitotic chromosome scaffolds. *J Cell Biol*. 100:1706-15.

Eivazova, E.R., Y.S. Vassetzky, and T.M. Aune. 2007. Selective matrix attachment regions in T helper cell subsets support loop conformation in the Ifng gene. *Genes Immun*. 8:35-43.

Farache, G., S.V. Razin, F.R. Targa, and K. Scherrer. 1990. Organization of the 3'-boundary of the chicken alpha globin gene domain and characterization of a CR 1-specific protein binding site. *Nucleic Acids Res*. 18:401-9.

Fernando, M.M., C.R. Stevens, E.C. Walsh, P.L. De Jager, P. Goyette, R.M. Plenge, T.J. Vyse, and J.D. Rioux. 2008. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet*. 4:e1000024.

Fey, E.G., and S. Penman. 1988. Nuclear matrix proteins reflect cell type of origin in cultured human cells. *Proc Natl Acad Sci U S A*. 85:121-5.

Fiegler, H., P. Carr, E.J. Douglas, D.C. Burford, S. Hunt, C.E. Scott, J. Smith, D. Vetrie, P. Gorman, I.P. Tomlinson, and N.P. Carter. 2003. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer*. 36:361-74.

Fillingham, J., M.C. Keogh, and N.J. Krogan. 2006. GammaH2AX and its role in DNA double-strand break repair. *Biochem Cell Biol*. 84:568-77.

Finch, J.T., and A. Klug. 1976. Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*. 73:1897-901.

Finlan, L.E., D. Sproul, I. Thomson, S. Boyle, E. Kerr, P. Perry, B. Ylstra, J.R. Chubb, and W.A. Bickmore. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*. 4:e1000039.

Fiorini, A., S. Gouveia Fde, and M.A. Fernandez. 2006. Scaffold/Matrix Attachment Regions and intrinsic DNA curvature. *Biochemistry (Mosc)*. 71:481-8.

Fischle, W., B.S. Tseng, H.L. Dormann, B.M. Ueberheide, B.A. Garcia, J. Shabanowitz, D.F. Hunt, H. Funabiki, and C.D. Allis. 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*. 438:1116-22.

Frisch, M., K. Frech, A. Klingenhoff, K. Cartharius, I. Liebich, and T. Werner. 2002. In silico prediction of scaffold/matrix attachment regions in large genomic sequences. *Genome Res*. 12:349-54.

Galande, S., P.K. Purbey, D. Notani, and P.P. Kumar. 2007. The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev*. 17:408-14.

Gall, J.G. 2000. Cajal bodies: the first 100 years. *Annu Rev Cell Dev Biol*. 16:273-300.

Gambelunghe, G., R. Gerli, E.B. Bocci, P. Del Sindaco, M. Ghaderi, C.B. Sanjeevi, O. Bistoni, V. Bini, and A. Falorni. 2005. Contribution of MHC class I chain-related A (MICA) gene polymorphism to genetic susceptibility for systemic lupus erythematosus. *Rheumatology (Oxford)*. 44:287-92.

Gerdes, M.G., K.C. Carter, P.T. Moen, Jr., and J.B. Lawrence. 1994. Dynamic changes in the higher-level chromatin organization of specific sequences revealed by in situ hybridization to nuclear halos. *J Cell Biol*. 126:289-304.

Gerner, C., K. Holzmann, R. Grimm, and G. Sauermann. 1998. Similarity between nuclear matrix proteins of various cells revealed by an improved isolation method. *J Cell Biochem*. 71:363-74.

Geyer, P.K. 1997. The role of insulator elements in defining domains of gene expression. *Curr Opin Genet Dev*. 7:242-8.

Gilbert, N., S. Boyle, H. Fiegler, K. Woodfine, N.P. Carter, and W.A. Bickmore. 2004. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*. 118:555-66.

Girard-Reydet, C., D. Gregoire, Y. Vassetzky, and M. Mechali. 2004. DNA replication initiates at domains overlapping with nuclear matrix attachment regions in the xenopus and mouse c-myc promoter. *Gene*. 332:129-38.

Girod, P.A., D.Q. Nguyen, D. Calabrese, S. Puttini, M. Grandjean, D. Martinet, A. Regamey, D. Saugy, J.S. Beckmann, P. Bucher, and N. Mermod. 2007. Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat Methods*.

Girod, P.A., M. Zahn-Zabal, and N. Mermod. 2005. Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnol Bioeng*. 91:1-11.

Glazko, G.V., E.V. Koonin, I.B. Rogozin, and S.A. Shabalina. 2003. A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet*. 19:119-24.

Glazko, G.V., I.B. Rogozin, and M.V. Glazkov. 2001. Comparative sudy and prediction of DNA fragments associated with various elements of the nuclear matrix. *Biochimica et Biophysica Acta*. 1517:351-364.

Gobin, S.J., V. Keijsers, M. van Zutphen, and P.J. van den Elsen. 1998. The role of enhancer A in the locus-specific transactivation of classical and nonclassical HLA class I genes by nuclear factor kappa B. *J Immunol*. 161:2276-83.

Gobin, S.J., A. Peijnenburg, V. Keijsers, and P.J. van den Elsen. 1997. Site alpha is crucial for two routes of IFN gamma-induced MHC class I transactivation: the ISRE-mediated route and a novel pathway involving CIITA. *Immunity*. 6:601-11.

Gobin, S.J., and P.J. van den Elsen. 1999. The regulation of HLA class I expression: is HLA-G the odd one out? *Semin Cancer Biol*. 9:55-9.

Gobin, S.J., and P.J. van den Elsen. 2000. Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F, and HLA-G. *Hum Immunol*. 61:1102-7.

Gobin, S.J., M. van Zutphen, A.M. Woltman, and P.J. van den Elsen. 1999. Transactivation of classical and nonclassical HLA class I genes through the IFN-stimulated response element. *J Immunol*. 163:1428-34.

Goldman, R.D., Y. Gruenbaum, R.D. Moir, D.K. Shumaker, and T.P. Spann. 2002. Nuclear lamins: building blocks of nuclear architecture. *Genes Dev*. 16:533-47.

Gorski, S., and T. Misteli. 2005. Systems biology in the cell nucleus. *J Cell Sci*. 118:4083-92.

Grande, M.A., I. van der Kraan, L. de Jong, and R. van Driel. 1997. Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J Cell Sci*. 110 (Pt 15):1781-91.

Gregory, S.G., K.F. Barlow, K.E. McLay, R. Kaul, D. Swarbreck, A. Dunham, C.E. Scott, K.L. Howe, K. Woodfine, C.C. Spencer, M.C. Jones, C. Gillson, S. Searle, Y. Zhou, F. Kokocinski, L. McDonald, R. Evans, K. Phillips, A. Atkinson, R. Cooper, C. Jones, R.E. Hall, T.D. Andrews, C. Lloyd, R. Ainscough, J.P. Almeida, K.D. Ambrose, F. Anderson, R.W. Andrew, R.I. Ashwell, K. Aubin, A.K. Babbage, C.L. Bagguley, J. Bailey, H. Beasley, G. Bethel, C.P. Bird, S. Bray-Allen, J.Y. Brown, A.J. Brown, D. Buckley, J. Burton, J. Bye, C. Carder, J.C. Chapman, S.Y. Clark, G. Clarke, C. Clee, V. Cobley, R.E. Collier, N. Corby, G.J. Coville, J. Davies, R. Deadman, M. Dunn, M. Earthrowl, A.G. Ellington, H. Errington, A. Frankish, J. Frankland, L. French, P. Garner, J. Garnett, L. Gay, M.R. Ghori, R. Gibson, L.M. Gilby, W. Gillett, R.J. Glithero, D.V. Grafham, C. Griffiths, S. Griffiths-Jones, R. Grocock, S. Hammond, E.S. Harrison, E. Hart, E. Haugen, P.D. Heath, S. Holmes, K. Holt, P.J. Howden, A.R. Hunt, S.E. Hunt, G. Hunter, J. Isherwood, R. James, C. Johnson, D. Johnson, A. Joy, M. Kay, J.K. Kershaw, M. Kibukawa, A.M. Kimberley, A. King, A.J. Knights, H. Lad, G. Laird, S. Lawlor, D.A. Leongamornlert, D.M. Lloyd, et al. 2006. The DNA sequence and biological annotation of human chromosome 1. *Nature*. 441:315-21.

Grewal, S.I., and S.C. Elgin. 2002. Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Genet Dev*. 12:178-87.
Gribbon, C., R. Dahm, A.R. Prescott, and R.A. Quinlan. 2002. Association of the nuclear matrix component NuMA with the Cajal body and nuclear speckle compartments during transitions in transcriptional activity in lens cell differentiation. *Eur J Cell Biol*. 81:557-66.
Gromme, M., F.G. Uytdehaag, H. Janssen, J. Calafat, R.S. van Binnendijk, M.J. Kenter, A. Tulp, D. Verwoerd, and J. Neefjes. 1999. Recycling MHC class I molecules and endosomal peptide loading. *Proc Natl Acad Sci U S A*. 96:10326-31.
Grunstein, M. 1997. Histone acetylation in chromatin structure and transcription. *Nature*. 389:349-52.
Halweg, C., W.F. Thompson, and S. Spiker. 2005. The rb7 matrix attachment region increases the likelihood and magnitude of transgene expression in tobacco cells: a flow cytometric study. *Plant Cell*. 17:418-29.
Hancock, R. 2004a. A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *J Struct Biol*. 146:281-90.
Hancock, R. 2004b. Internal organisation of the nucleus: assembly of compartments by macromolecular crowding and the nuclear matrix model. *Biol Cell*. 96:595-601.
Hart, C.M., and U.K. Laemmli. 1998. Facilitation of chromatin dynamics by SARs. *Curr Opin Genet Dev*. 8:519-25.
Harton, J.A., and J.P. Ting. 2000. Class II transactivator: mastering the art of major histocompatibility complex expression. *Mol Cell Biol*. 20:6185-94.
Hauptmann, G., and S. Bahram. 2004. Genetics of the central MHC. *Curr Opin Immunol*. 16:668-72.
Heinzelmann-Schwarz, V.A., M. Gardiner-Garden, S.M. Henshall, J. Scurry, R.A. Scolyer, M.J. Davies, M. Heinzelmann, L.H. Kalish, A. Bali, J.G. Kench, L.S. Edwards, P.M. Vanden Bergh, N.F. Hacker, R.L. Sutherland, and P.M. O'Brien. 2004. Overexpression of the cell adhesion molecules DDR1, Claudin 3, and Ep-CAM in metaplastic ovarian epithelium and ovarian cancer. *Clin Cancer Res*. 10:4427-36.
Heng, H.H., S. Goetze, C.J. Ye, G. Liu, J.B. Stevens, S.W. Bremer, S.M. Wykes, J. Bode, and S.A. Krawetz. 2004. Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J Cell Sci*. 117:999-1008.
Heng, H.H., S.A. Krawetz, W. Lu, S. Bremer, G. Liu, and C.J. Ye. 2001. Re-defining the chromatin loop domain. *Cytogenet Cell Genet*. 93:155-61.
Herrmann, H., and R. Foisner. 2003. Intermediate filaments: novel assembly models and exciting new functions for nuclear lamins. *Cell Mol Life Sci*. 60:1607-12.
Hibino, Y., C.S. Kumar, T.M. Mariano, D.H. Lai, and S. Pestka. 1992. Chimeric interferon-gamma receptors demonstrate that an accessory factor required for activity interacts with the extracellular domain. *J Biol Chem*. 267:3741-9.
Hill, A.V., C.E. Allsopp, D. Kwiatkowski, N.M. Anstey, P. Twumasi, P.A. Rowe, S. Bennett, D. Brewster, A.J. McMichael, and B.M. Greenwood. 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature*. 352:595-600.
Hoare, H.L., L.C. Sullivan, G. Pietra, C.S. Clements, E.J. Lee, L.K. Ely, T. Beddoe, M. Falco, L. Kjer-Nielsen, H.H. Reid, J. McCluskey, L. Moretta, J. Rossjohn, and A.G. Brooks. 2006. Structural basis for a major histocompatibility complex class Ib-restricted T cell response. *Nat Immunol*. 7:256-64.
Holmskov, U., S. Thiel, and J.C. Jensenius. 2003. Collections and ficolins: humoral lectins of the innate immune defense. *Annu Rev Immunol*. 21:547-78.
Horton, R., L. Wilming, V. Rand, R.C. Lovering, E.A. Bruford, V.K. Khodiyar, M.J. Lush, S. Povey, C.C. Talbot, Jr., M.W. Wright, H.M. Wain, J. Trowsdale, A.

Ziegler, and S. Beck. 2004. Gene map of the extended human MHC. *Nat Rev Genet*. 5:889-99.
Howman, E.V., K.J. Fowler, A.J. Newson, S. Redward, A.C. MacDonald, P. Kalitsis, and K.H. Choo. 2000. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci U S A*. 97:1148-53.
Hozak, P., A.B. Hassan, D.A. Jackson, and P.R. Cook. 1993. Visualization of replication factories attached to nucleoskeleton. *Cell*. 73:361-73.
Hubbard, T.J., B.L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S.C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X.M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. 2007. Ensembl 2007. *Nucleic Acids Res*. 35:D610-7.
Hviid, T.V. 2006. HLA-G in human reproduction: aspects of genetics, function and pregnancy complications. *Hum Reprod Update*. 12:209-32.
Iarovaia, O.V., P. Shkumatov, and S.V. Razin. 2004. Breakpoint cluster regions of the AML-1 and ETO genes contain MAR elements and are preferentially associated with the nuclear matrix in proliferating HEL cells. *J Cell Sci*. 117:4583-90.
Iborra, F.J., A. Pombo, D.A. Jackson, and P.R. Cook. 1996. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci*. 109 (Pt 6):1427-36.
Iizuka, M., T. Matsui, H. Takisawa, and M.M. Smith. 2006. Regulation of replication licensing by acetyltransferase Hbo1. *Mol Cell Biol*. 26:1098-108.
Ioudinkova, E., A. Petrov, S.V. Razin, and Y.S. Vassetzky. 2005. Mapping long-range chromatin organization within the chicken alpha-globin gene domain using oligonucleotide DNA arrays. *Genomics*. 85:143-51.
Ishii, K., G. Arib, C. Lin, G. Van Houwe, and U.K. Laemmli. 2002. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell*. 109:551-62.
Ivanova, M., K.M. Dobrzycka, S. Jiang, K. Michaelis, R. Meyer, K. Kang, B. Adkins, O.A. Barski, S. Zubairy, J. Divisova, A.V. Lee, and S. Oesterreich. 2005. Scaffold attachment factor B1 functions in development, growth, and reproduction. *Mol Cell Biol*. 25:2995-3006.
Jackson, D.A. 2003. The principles of nuclear structure. *Chromosome Res*. 11:387-401.
Jackson, D.A., J. Bartlett, and P.R. Cook. 1996. Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: the role of transcription units. *Nucleic Acids Res*. 24:1212-9.
Jackson, D.A., and P.R. Cook. 1986. Replication occurs at a nucleoskeleton. *Embo J*. 5:1403-10.
Jackson, D.A., and P.R. Cook. 1988. Visualization of a filamentous nucleoskeleton with a 23 nm axial repeat. *Embo J*. 7:3667-77.
Jackson, D.A., P. Dickinson, and P.R. Cook. 1990. Attachment of DNA to the nucleoskeleton of HeLa cells examined using physiological conditions. *Nucleic Acids Res*. 18:4385-93.
Jackson, D.A., A. Dolle, G. Robertson, and P.R. Cook. 1992. The attachments of chromatin loops to the nucleoskeleton. *Cell Biol Int Rep*. 16:687-96.
Jackson, D.A., A.B. Hassan, R.J. Errington, and P.R. Cook. 1993. Visualization of focal sites of transcription within human nuclei. *Embo J*. 12:1059-65.

Jackson, D.A., and A. Pombo. 1998. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol*. 140:1285-95.

Jackson, J.A., A.V. Trevino, M.C. Herzig, T.S. Herman, and J.M. Woynarowski. 2003. Matrix attachment region (MAR) properties and abnormal expansion of AT island minisatellites in FRA16B fragile sites in leukemic CEM cells. *Nucleic Acids Res*. 31:6354-64.

Jacobs, S.A., S.D. Taverna, Y. Zhang, S.D. Briggs, J. Li, J.C. Eissenberg, C.D. Allis, and S. Khorasanizadeh. 2001. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *Embo J*. 20:5232-41.

Johnson, C.N., and L.S. Levy. 2005. Matrix attachment regions as targets for retroviral integration. *Virol J*. 2:68.

Jones, E.Y., L. Fugger, J.L. Strominger, and C. Siebold. 2006. MHC class II proteins and disease: a structural perspective. *Nat Rev Immunol*. 6:271-82.

Jordan, I.K., I.B. Rogozin, G.V. Glazko, and E.V. Koonin. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* 19:68-72.

Kanda, T., K.F. Sullivan, and G.M. Wahl. 1998. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr Biol*. 8:377-85.

Kasahara, M., M.F. Flajnik, T. Ishibashi, and T. Natori. 1995. Evolution of the major histocompatibility complex: a current overview. *Transpl Immunol*. 3:1-20.

Kennedy, B.K., D.A. Barbie, M. Classon, N. Dyson, and E. Harlow. 2000. Nuclear organization of DNA replication in primary mammalian cells. *Genes Dev*. 14:2855-68.

Khorasanizadeh, S. 2004. The nucleosome: from genomic organization to genomic regulation. *Cell*. 116:259-72.

Kieffer, L.J., J.M. Greally, I. Landres, S. Nag, Y. Nakajima, T. Kohwi-Shigematsu, and P.B. Kavathas. 2002. Identification of a candidate regulatory region in the human CD8 gene complex by colocalization of DNase I hypersensitive sites and matrix attachment regions which bind SATB1 and GATA-3. *J Immunol*. 168:3915-22.

Kim, T.H., Z.K. Abdullaev, A.D. Smith, K.A. Ching, D.I. Loukinov, R.D. Green, M.Q. Zhang, V.V. Lobanenkov, and B. Ren. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*. 128:1231-45.

Kimura, H., Y. Tao, R.G. Roeder, and P.R. Cook. 1999. Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Mol Cell Biol*. 19:5383-92.

Kioussis, D. 2005. Gene regulation: kissing chromosomes. *Nature*. 435:579-80.

Kipp, M., F. Gohring, T. Ostendorp, C.M. van Drunen, R. van Driel, M. Przybylski, and F.O. Fackelmayer. 2000. SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol Cell Biol*. 20:7480-9.

Kireeva, N., M. Lakonishok, I. Kireev, T. Hirano, and A.S. Belmont. 2004. Visualization of early chromosome condensation: a hierarchical folding, axial glue model of chromosome structure. *J Cell Biol*. 166:775-85.

Kiriakidou, M., D.A. Driscoll, J.M. Lopez-Guisa, and J.F. Strauss, 3rd. 1997. Cloning and expression of primate Daxx cDNAs and mapping of the human gene to chromosome 6p21.3 in the MHC region. *DNA Cell Biol*. 16:1289-98.

Kitamura, E., J.J. Blow, and T.U. Tanaka. 2006. Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell*. 125:1297-308.

Klein, J., and A. Sato. 1998. Birth of the major histocompatibility complex. *Scand J Immunol*. 47:199-209.
Kloetzel, P.M., and F. Ossendorp. 2004. Proteasome and peptidase function in MHC-class-I-mediated antigen presentation. *Curr Opin Immunol*. 16:76-81.
Kohwi-Shigematsu, T., and Y. Kohwi. 1990. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry*. 29:9551-60.
Kosak, S.T., and M. Groudine. 2004. Form follows function: The genomic organization of cellular differentiation. *Genes Dev*. 18:1371-84.
Kosak, S.T., J.A. Skok, K.L. Medina, R. Riblet, M.M. Le Beau, A.G. Fisher, and H. Singh. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*. 296:158-62.
Kouzarides, T. 2007. Chromatin modifications and their function. *Cell*. 128:693-705.
Kramer, J.A., and S.A. Krawetz. 1996. Nuclear matrix interactions within the sperm genome. *J Biol Chem*. 271:11619-22.
Kramer, J.A., and S.A. Krawetz. 1997. PCR-based assay to determine nuclear matrix association. *Biotechniques*. 22:826-8.
Krawczyk, M., N. Peyraud, N. Rybtsova, K. Masternak, P. Bucher, E. Barras, and W. Reith. 2004. Long distance control of MHC class II expression by multiple distal enhancers regulated by regulatory factor X complex and CIITA. *J Immunol*. 173:6200-10.
Krawczyk, M., Q. Seguin-Estevez, E. Leimgruber, P. Sperisen, C. Schmid, P. Bucher, and W. Reith. 2008. Identification of CIITA regulated genetic module dedicated for antigen presentation. *PLoS Genet*. 4:e1000058.
Krawetz, S.A., S. Draghici, R. Goodrich, Z. Liu, and G.C. Ostermeier. 2005. In silico and wet-bench identification of nuclear matrix attachment regions. *Methods Mol Med*. 108:439-58.
Kulkarni, A., L. Pavithra, S. Rampalli, D. Mogare, K. Babu, G. Shiekh, S. Ghosh, and S. Chattopadhyay. 2004. HIV-1 integration sites are flanked by potential MARs that alone can act as promoters. *Biochem Biophys Res Commun*. 322:672-7.
Kumanovics, A., T. Takada, and K.F. Lindahl. 2003. Genomic organization of the mammalian MHC. *Annu Rev Immunol*. 21:629-57.
Kumar, P.P., O. Bischof, P.K. Purbey, D. Notani, H. Urlaub, A. Dejean, and S. Galande. 2007a. Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol*. 9:45-56.
Kumar, P.P., S. Mehta, P.K. Purbey, D. Notani, R.S. Jayani, H.J. Purohit, D.V. Raje, D.S. Ravi, R.R. Bhonde, D. Mitra, and S. Galande. 2007b. SATB1-binding sequences and Alu-like motifs define a unique chromatin context in the vicinity of human immunodeficiency virus type 1 integration sites. *J Virol*. 81:5617-27.
Kumaran, R.I., R. Thakar, and D.L. Spector. 2008. Chromatin dynamics and gene positioning. *Cell*. 132:929-34.
Kurdistani, S.K., and M. Grunstein. 2003. Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol*. 4:276-84.
Kurdistani, S.K., S. Tavazoie, and M. Grunstein. 2004. Mapping global histone acetylation patterns to gene expression. *Cell*. 117:721-33.
Lagarkova, M.A., E. Svetlova, M. Giacca, A. Falaschi, and S.V. Razin. 1998. DNA loop anchorage region colocalizes with the replication origin located downstream to the human gene encoding lamin B2. *J Cell Biochem*. 69:13-8.
Lamond, A.I., and D.L. Spector. 2003. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol*. 4:605-12.

Law, S.K., and A.W. Dodds. 1997. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci*. 6:263-74.
Lee, K.K., and J.L. Workman. 2007. Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol*. 8:284-95.
Lee, Y.J., and E.N. Benveniste. 1996. Stat1 alpha expression is involved in IFN-gamma induction of the class II transactivator and class II MHC genes. *J Immunol*. 157:1559-68.
Lepin, E.J., J.M. Bastin, D.S. Allan, G. Roncador, V.M. Braud, D.Y. Mason, P.A. van der Merwe, A.J. McMichael, J.I. Bell, S.H. Powis, and C.A. O'Callaghan. 2000. Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors. *Eur J Immunol*. 30:3552-61.
Li, B., M. Carey, and J.L. Workman. 2007. The role of chromatin during transcription. *Cell*. 128:707-19.
Li, Y., G.D. Kao, B.A. Garcia, J. Shabanowitz, D.F. Hunt, J. Qin, C. Phelan, and M.A. Lazar. 2006. A novel histone deacetylase pathway regulates mitosis by modulating Aurora B kinase activity. *Genes Dev*. 20:2566-79.
Liang, G., J.C. Lin, V. Wei, C. Yoo, J.C. Cheng, C.T. Nguyen, D.J. Weisenberger, G. Egger, D. Takai, F.A. Gonzales, and P.A. Jones. 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U S A*. 101:7357-62.
Liebich, I., J. Bode, I. Reuter, and E. Wingender. 2002. Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res*. 30:3433-42.
Lin, W.J., J.D. Gary, M.C. Yang, S. Clarke, and H.R. Herschman. 1996. The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J Biol Chem*. 271:15034-44.
Linnemann, A.K., A.E. Platts, N. Doggett, A. Gluch, J. Bode, and S.A. Krawetz. 2007. Genomewide identification of nuclear matrix attachment regions: an analysis of methods. *Biochem Soc Trans*. 35:612-7.
Litt, M.D., M. Simpson, M. Gaszner, C.D. Allis, and G. Felsenfeld. 2001. Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science*. 293:2453-5.
Liu, B., J. Wang, K.M. Chan, W.M. Tjia, W. Deng, X. Guan, J.D. Huang, K.M. Li, P.Y. Chau, D.J. Chen, D. Pei, A.M. Pendas, J. Cadinanos, C. Lopez-Otin, H.F. Tse, C. Hutchison, J. Chen, Y. Cao, K.S. Cheah, K. Tryggvason, and Z. Zhou. 2005. Genomic instability in laminopathy-based premature aging. *Nat Med*. 11:780-5.
Loots, G.G., R.M. Locksley, C.M. Blankespoor, Z.E. Wang, W. Miller, E.M. Rubin, and K.A. Frazer. 2000. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*. 288:136-40.
Louis-Plence, P., C.S. Moreno, and J.M. Boss. 1997. Formation of a regulatory factor X/X2 box-binding protein/nuclear factor-Y multiprotein complex on the conserved regulatory regions of HLA class II genes. *J Immunol*. 159:3899-909.
Luderus, M.E., B. van Steensel, L. Chong, O.C. Sibon, F.F. Cremers, and T. de Lange. 1996. Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *J Cell Biol*. 135:867-81.
Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 389:251-60.
Lukacs, G.L., P. Haggie, O. Seksek, D. Lechardeur, N. Freedman, and A.S. Verkman. 2000. Size-dependent DNA mobility in cytoplasm and nucleus. *J Biol Chem*. 275:1625-9.

Lukasova, E., S. Kozubek, M. Kozubek, J. Kjeronska, L. Ryznar, J. Horakova, E. Krahulcova, and G. Horneck. 1997. Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. *Hum Genet*. 100:525-35.

Ma, H., J. Samarabandu, R.S. Devdhar, R. Acharya, P.C. Cheng, C. Meng, and R. Berezney. 1998. Spatial and temporal dynamics of DNA replication sites in mammalian cells. *J Cell Biol*. 143:1415-25.

Maas, N.L., K.M. Miller, L.G. DeFazio, and D.P. Toczyski. 2006. Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. *Mol Cell*. 23:109-19.

Maeshima, K., and U.K. Laemmli. 2003. A two-step scaffolding model for mitotic chromosome assembly. *Dev Cell*. 4:467-80.

Mahy, N.L., P.E. Perry, and W.A. Bickmore. 2002. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol*. 159:753-63.

Makalowski, W. 2001. The human genome structure and organization. *Acta Biochim Pol*. 48:587-98.

Marko, J.F. 2008. Micromechanical studies of mitotic chromosomes. *Chromosome Res*. 16:469-97.

Marsden, M.P., and U.K. Laemmli. 1979. Metaphase chromosome structure: evidence for a radial loop model. *Cell*. 17:849-58.

Martens, J.H., M. Verlaan, E. Kalkhoven, J.C. Dorsman, and A. Zantema. 2002. Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. *Mol Cell Biol*. 22:2598-606.

Martin, B.K., K.C. Chin, J.C. Olsen, C.A. Skinner, A. Dey, K. Ozato, and J.P. Ting. 1997. Induction of MHC class I expression by the MHC class II transactivator CIITA. *Immunity*. 6:591-600.

Martins, R.P., and S.A. Krawetz. 2007. Decondensing the protamine domain for transcription. *Proc Natl Acad Sci U S A*. 104:8340-5.

Masternak, K., A. Muhlethaler-Mottet, J. Villard, M. Zufferey, V. Steimle, and W. Reith. 2000. CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex. *Genes Dev*. 14:1156-66.

Masternak, K., N. Peyraud, M. Krawczyk, E. Barras, and W. Reith. 2003. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol*. 4:132-7.

Matsuzaka, Y., K. Tounai, A. Denda, M. Tomizawa, S. Makino, K. Okamoto, N. Keicho, A. Oka, J.K. Kulski, G. Tamiya, and H. Inoko. 2002. Identification of novel candidate genes in the diffuse panbronchiolitis critical region of the class I human MHC. *Immunogenetics*. 54:301-9.

Mattern, K.A., B.M. Humbel, A.O. Muijsers, L. de Jong, and R. van Driel. 1996. hnRNP proteins and B23 are the major proteins of the internal nuclear matrix of HeLa S3 cells. *J Cell Biochem*. 62:275-89.

Mattern, K.A., R.E. van Goethem, L. de Jong, and R. van Driel. 1997. Major internal nuclear matrix proteins are common to different human cell types. *J Cell Biochem*. 65:42-52.

McKnight, R.A., A. Shamay, L. Sankaran, R.J. Wall, and L. Hennighausen. 1992. Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proc Natl Acad Sci U S A*. 89:6943-7.

Meaburn, K.J., E. Cabuy, G. Bonne, N. Levy, G.E. Morris, G. Novelli, I.R. Kill, and J.M. Bridger. 2007. Primary laminopathy fibroblasts display altered genome organization and apoptosis. *Aging Cell*. 6:139-53.

Meneghini, M.D., M. Wu, and H.D. Madhani. 2003. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell*. 112:725-36.

Meroni, G., and G. Diez-Roux. 2005. TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *Bioessays*. 27:1147-57.

Mesner, L.D., J.L. Hamlin, and P.A. Dijkwel. 2003. The matrix attachment region in the Chinese hamster dihydrofolate reductase origin of replication may be required for local chromatid separation. *Proc Natl Acad Sci U S A*. 100:3281-6.

Michalowski, S.M., G.C. Allen, G.E. Hall, Jr., W.F. Thompson, and S. Spiker. 1999. Characterization of randomly-obtained matrix attachment regions (MARs) from higher plants. *Biochemistry*. 38:12795-804.

Mirkovitch, J., M.E. Mirault, and U.K. Laemmli. 1984. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*. 39:223-32.

Misteli, T. 2001. The concept of self-organization in cellular architecture. *J Cell Biol*. 155:181-5.

Misteli, T. 2004. Spatial positioning; a new dimension in genome function. *Cell*. 119:153-6.

Misteli, T. 2005. Concepts in nuclear architecture. *Bioessays*. 27:477-87.

Misteli, T. 2007. Beyond the sequence: cellular organization of genome function. *Cell*. 128:787-800.

Mitchell, J.A., and P. Fraser. 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev*. 22:20-5.

Moir, R.D., M. Montag-Lowy, and R.D. Goldman. 1994. Dynamic properties of nuclear lamins: lamin B is associated with sites of DNA replication. *J Cell Biol*. 125:1201-12.

Monzo, M., R. Rosell, J.J. Sanchez, J.S. Lee, A. O'Brate, J.L. Gonzalez-Larriba, V. Alberola, J.C. Lorenzo, L. Nunez, J.Y. Ro, and C. Martin. 1999. Paclitaxel resistance in non-small-cell lung cancer associated with beta-tubulin gene mutations. *J Clin Oncol*. 17:1786-93.

Morgan, B.P., and C.L. Harris. 2003. Complement therapeutics; history and current progress. *Mol Immunol*. 40:159-70.

Muchir, A., B.G. van Engelen, M. Lammens, J.M. Mislow, E. McNally, K. Schwartz, and G. Bonne. 2003. Nuclear envelope alterations in fibroblasts from LGMD1B patients carrying nonsense Y259X heterozygous or homozygous mutation in lamin A/C gene. *Exp Cell Res*. 291:352-62.

Mudhasani, R., and J.D. Fontes. 2002. The class II transactivator requires brahma-related gene 1 to activate transcription of major histocompatibility complex class II genes. *Mol Cell Biol*. 22:5019-26.

Mulholland, P.J., H. Fiegler, C. Mazzanti, P. Gorman, P. Sasieni, J. Adams, T.A. Jones, J.W. Babbage, R. Vatcheva, K. Ichimura, P. East, C. Poullikas, V.P. Collins, N.P. Carter, I.P. Tomlinson, and D. Sheer. 2006. Genomic profiling identifies discrete deletions associated with translocations in glioblastoma multiforme. *Cell Cycle*. 5:783-91.

Munkel, C., R. Eils, S. Dietzel, D. Zink, C. Mehring, G. Wedemann, T. Cremer, and J. Langowski. 1999. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol*. 285:1053-65.

Murmann, A.E., J. Gao, M. Encinosa, M. Gautier, M.E. Peter, R. Eils, P. Lichter, and J.D. Rowley. 2005. Local gene density predicts the spatial position of genetic loci in the interphase nucleus. *Exp Cell Res*. 311:14-26.

Muromoto, R., K. Sugiyama, A. Takachi, S. Imoto, N. Sato, T. Yamamoto, K. Oritani, K. Shimoda, and T. Matsuda. 2004. Physical and functional interactions between Daxx and DNA methyltransferase 1-associated protein, DMAP1. *J Immunol*. 172:2985-93.

Nabirochkin, S., M. Ossokina, and T. Heidmann. 1998. A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J Biol Chem*. 273:2473-9.

Nakayasu, H., and R. Berezney. 1989. Mapping replicational sites in the eucaryotic cell nucleus. *J Cell Biol*. 108:1-11.

Namciu, S.J., K.B. Blochlinger, and R.E. Fournier. 1998. Human matrix attachment regions insulate transgene expression from chromosomal position effects in Drosophila melanogaster. *Mol Cell Biol*. 18:2382-91.

Nayler, O., W. Stratling, J.P. Bourquin, I. Stagljar, L. Lindemann, H. Jasper, A.M. Hartmann, F.O. Fackelmayer, A. Ullrich, and S. Stamm. 1998. SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res*. 26:3542-9.

Neves, H., C. Ramos, M.G. da Silva, A. Parreira, and L. Parreira. 1999. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood*. 93:1197-207.

Nickerson, J. 2001. Experimental observations of a nuclear matrix. *J Cell Sci*. 114:463-74.

Nickerson, J.A., B.J. Blencowe, and S. Penman. 1995. The architectural organization of nuclear metabolism. *Int Rev Cytol*. 162A:67-123.

Niimura, Y., and T. Gojobori. 2002. In silico chromosome staining: reconstruction of Giemsa bands from the whole human genome sequence. *Proc Natl Acad Sci U S A*. 99:797-802.

Nikolova, V., C. Leimena, A.C. McMahon, J.C. Tan, S. Chandar, D. Jogia, S.H. Kesteven, J. Michalicek, R. Otway, F. Verheyen, S. Rainer, C.L. Stewart, D. Martin, M.P. Feneley, and D. Fatkin. 2004. Defects in nuclear structure and function promote dilated cardiomyopathy in lamin A/C-deficient mice. *J Clin Invest*. 113:357-69.

Noma, K., H.P. Cam, R.J. Maraia, and S.I. Grewal. 2006. A role for TFIIIC transcription factor complex in genome organization. *Cell*. 125:859-72.

Nowak, S.J., and V.G. Corces. 2004. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet*. 20:214-20.

O'Sullivan, J.M., S.M. Tan-Wong, A. Morillon, B. Lee, J. Coles, J. Mellor, and N.J. Proudfoot. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*. 36:1014-8.

Oegema, K., W.F. Marshall, J.W. Sedat, and B.M. Alberts. 1997. Two proteins that cycle asynchronously between centrosomes and nuclear structures: Drosophila CP60 and CP190. *J Cell Sci*. 110 (Pt 14):1573-83.

Oesterreich, S. 2003. Scaffold attachment factors SAFB1 and SAFB2: Innocent bystanders or critical players in breast tumorigenesis? *J Cell Biochem*. 90:653-61.

Ogg, S.C., and A.I. Lamond. 2002. Cajal bodies and coilin--moving towards function. *J Cell Biol*. 159:17-21.

Ohta, S., Y. Tatsumi, M. Fujita, T. Tsurimoto, and C. Obuse. 2003. The ORC1 cycle in human cells: II. Dynamic changes in the human ORC complex during the cell cycle. *J Biol Chem*. 278:41535-40.

Olson, M.O., M. Dundr, and A. Szebeni. 2000. The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol*. 10:189-96.

Osborne, C.S., L. Chakalova, K.E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J.A. Mitchell, S. Lopes, W. Reik, and P. Fraser. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*. 36:1065-71.

Osborne, C.S., L. Chakalova, J.A. Mitchell, A. Horton, A.L. Wood, D.J. Bolland, A.E. Corcoran, and P. Fraser. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol*. 5:e192.

Ostermeier, G.C., Z. Liu, R.P. Martins, R.R. Bharadwaj, J. Ellis, S. Draghici, and S.A. Krawetz. 2003. Nuclear matrix association of the human beta-globin locus utilizing a novel approach to quantitative real-time PCR. *Nucleic Acids Res*. 31:3257-66.

Ottaviani, D., E. Lever, R. Mitter, T. Jones, T. Forshew, R. Christova, E.M. Tomazou, V.K. Rakyan, S.A. Krawetz, E.P. Platts, B. Segarane, S. Beck, and D. Sheer. 2008a. Reconfiguration of genomic anchors upon transcriptional activation of the human Major Histocompatibility Complex. *Genome Res*.(in press).

Ottaviani, D., E. Lever, P. Takousis, and D. Sheer. 2008b. Anchoring the genome. *Genome Biol*. 9:201.

Parada, L.A., S. Sotiriou, and T. Misteli. 2004. Spatial genome organization. *Exp Cell Res*. 296:64-70.

Patturajan, M., X. Wei, R. Berezney, and J.L. Corden. 1998. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. *Mol Cell Biol*. 18:2406-15.

Paulson, J.R., and U.K. Laemmli. 1977. The structure of histone-depleted metaphase chromosomes. *Cell*. 12:817-28.

Payne, R., M. Tripp, J. Weigle, W. Bodmer, and J. Bodmer. 1964. A New Leukocyte Isoantigen System in Man. *Cold Spring Harb Symp Quant Biol*. 29:285-95.

Pederson, T. 1998. Thinking about a nuclear matrix. *J Mol Biol*. 277:147-59.

Pederson, T. 2000. Half a century of "the nuclear matrix". *Mol Biol Cell*. 11:799-805.

Pende, D., S. Parolini, A. Pessino, S. Sivori, R. Augugliaro, L. Morelli, E. Marcenaro, L. Accame, A. Malaspina, R. Biassoni, C. Bottino, L. Moretta, and A. Moretta. 1999. Identification and molecular characterization of NKp30, a novel triggering receptor involved in natural cytotoxicity mediated by human natural killer cells. *J Exp Med*. 190:1505-16.

Petrov, A., I. Pirozhkova, G. Carnac, D. Laoudj, M. Lipinski, and Y.S. Vassetzky. 2006. Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. *Proc Natl Acad Sci U S A*. 103:6982-7.

Phi-van, L., C. Sellke, A. von Bodenhausen, and W.H. Stratling. 1998. An initiation zone of chromosomal DNA replication at the chicken lysozyme gene locus. *J Biol Chem*. 273:18300-7.

Pokholok, D.K., C.T. Harbison, S. Levine, M. Cole, N.M. Hannett, T.I. Lee, G.W. Bell, K. Walker, P.A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D.K. Gifford, and R.A. Young. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*. 122:517-27.

Pombo, A., and M.R. Branco. 2007. Functional organisation of the genome during interphase. *Curr Opin Genet Dev*. 17:451-5.

Pombo, A., D.A. Jackson, M. Hollinshead, Z. Wang, R.G. Roeder, and P.R. Cook. 1999. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *Embo J.* 18:2241-53.

Powis, S.H., I. Mockridge, A. Kelly, L.A. Kerr, R. Glynne, U. Gileadi, S. Beck, and J. Trowsdale. 1992. Polymorphism in a second ABC transporter gene located within the class II region of the human major histocompatibility complex. *Proc Natl Acad Sci U S A*. 89:1463-7.

Powis, S.H., S. Tonks, I. Mockridge, A.P. Kelly, J.G. Bodmer, and J. Trowsdale. 1993. Alleles and haplotypes of the MHC-encoded ABC transporters TAP1 and TAP2. *Immunogenetics*. 37:373-80.

Price, P., C. Witt, R. Allcock, D. Sayer, M. Garlepp, C.C. Kok, M. French, S. Mallal, and F. Christiansen. 1999. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol Rev*. 167:257-74.

Purbowasito, W., C. Suda, T. Yokomine, M. Zubair, T. Sado, K. Tsutsui, and H. Sasaki. 2004. Large-scale identification and mapping of nuclear matrix-attachment regions in the distal imprinted domain of mouse chromosome 7. *DNA Res*. 11:391-407.

Quina, A.S., M. Buschbeck, and L. Di Croce. 2006. Chromatin structure and epigenetics. *Biochem Pharmacol*. 72:1563-9.

Radichev, I., A. Parashkevova, and B. Anachkova. 2005. Initiation of DNA replication at a nuclear matrix-attached chromatin fraction. *J Cell Physiol*. 203:71-7.

Rajaiya, J., J.C. Nixon, N. Ayers, Z.P. Desgranges, A.L. Roy, and C.F. Webb. 2006. Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I. *Mol Cell Biol*. 26:4758-68.

Ramakrishnan, V. 1997. Histone structure and the organization of the nucleosome. *Annu Rev Biophys Biomol Struct*. 26:83-112.

Rampalli, S., A. Kulkarni, P. Kumar, D. Mogare, S. Galande, D. Mitra, and S. Chattopadhyay. 2003. Stimulation of Tat-independent transcriptional processivity from the HIV-1 LTR promoter by matrix attachment regions. *Nucleic Acids Res*. 31:3248-56.

Razin, S.V., Y.S. Vassetzky, and R. Hancock. 1991. Nuclear matrix attachment regions and topoisomerase II binding and reaction sites in the vicinity of a chicken DNA replication origin. *Biochem Biophys Res Commun*. 177:265-70.

Reddy, K.L., J.M. Zullo, E. Bertolino, and H. Singh. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*. 452:243-7.

Reith, W., and B. Mach. 2001. The bare lymphocyte syndrome and the regulation of MHC expression. *Annu Rev Immunol*. 19:331-73.

Reyes, J.C., C. Muchardt, and M. Yaniv. 1997. Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. *J Cell Biol*. 137:263-74.

Richards, E.J., and S.C. Elgin. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell*. 108:489-500.

Roh, T.Y., S. Cuddapah, and K. Zhao. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev*. 19:542-52.

Rohn, W.M., Y.J. Lee, and E.N. Benveniste. 1996. Regulation of class II MHC expression. *Crit Rev Immunol*. 16:311-30.

Roix, J.J., P.G. McQueen, P.J. Munson, L.A. Parada, and T. Misteli. 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet*. 34:287-91.

Rollinger-Holzinger, I., B. Eibl, M. Pauly, U. Griesser, F. Hentges, B. Auer, G. Pall, P. Schratzberger, D. Niederwieser, E.H. Weiss, and H. Zwierzina. 2000. LST1: a

gene with extensive alternative splicing and immunomodulatory function. *J Immunol*. 164:3169-76.
Rollini, P., S.J. Namciu, M.D. Marsden, and R.E. Fournier. 1999. Identification and characterization of nuclear matrix-attachment regions in the human serpin gene cluster at 14q32.1. *Nucleic Acids Res*. 27:3779-91.
Rubio, J.P., M. Bahlo, N. Tubridy, J. Stankovich, R. Burfoot, H. Butzkueven, C. Chapman, L. Johnson, M. Marriott, G. Mraz, B. Tait, C. Wilkinson, B. Taylor, T.P. Speed, S.J. Foote, and T.J. Kilpatrick. 2004. Extended haplotype analysis in the HLA complex reveals an increased frequency of the HFE-C282Y mutation in individuals with multiple sclerosis. *Hum Genet*. 114:573-80.
Sachs, R.K., G. van den Engh, B. Trask, H. Yokota, and J.E. Hearst. 1995. A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A*. 92:2710-4.
Saitoh, Y., and U.K. Laemmli. 1994. Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell*. 76:609-22.
Schneider, R., and R. Grosschedl. 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev*. 21:3027-43.
Schones, D.E., and K. Zhao. 2008. Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet*. 9:179-91.
Schrambach, S., M. Ardizzone, V. Leymarie, J. Sibilia, and S. Bahram. 2007. In vivo expression pattern of MICA and MICB and its relevance to auto-immunity and cancer. *PLoS ONE*. 2:e518.
Seo, J., M.M. Lozano, and J.P. Dudley. 2005. Nuclear matrix binding regulates SATB1-mediated transcriptional repression. *J Biol Chem*. 280:24600-9.
Shaposhnikov, S.A., S.B. Akopov, I.P. Chernov, P.D. Thomsen, C. Joergensen, A.R. Collins, E. Frengen, and L.G. Nikolaev. 2007. A map of nuclear matrix attachment regions within the breast cancer loss-of-heterozygosity region on human chromosome 16q22.1. *Genomics*. 89:354-61.
Shepherd, J.C., T.N. Schumacher, P.G. Ashton-Rickardt, S. Imaeda, H.L. Ploegh, C.A. Janeway, Jr., and S. Tonegawa. 1993. TAP1-dependent peptide translocation in vitro is ATP dependent and peptide selective. *Cell*. 74:577-84.
Shera, K.A., C.A. Shera, and J.K. McDougall. 2001. Small tumor virus genomes are integrated near nuclear matrix attachment regions in transformed cells. *J Virol*. 75:12339-46.
Shogren-Knaak, M., H. Ishii, J.M. Sun, M.J. Pazin, J.R. Davie, and C.L. Peterson. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*. 311:844-7.
Sim, R.B., and S.A. Tsiftsoglou. 2004. Proteases of the complement system. *Biochem Soc Trans*. 32:21-7.
Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 38:1348-54.
Singer, D.S., and J.E. Maguire. 1990. Regulation of the expression of class I MHC genes. *Crit Rev Immunol*. 10:235-57.
Singh, G.B., J.A. Kramer, and S.A. Krawetz. 1997. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res*. 25:1419-25.
Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 3:Article3.

Soloaga, A., S. Thomson, G.R. Wiggin, N. Rampersaud, M.H. Dyson, C.A. Hazzalin, L.C. Mahadevan, and J.S. Arthur. 2003. MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *Embo J*. 22:2788-97.

Spann, T.P., A.E. Goldman, C. Wang, S. Huang, and R.D. Goldman. 2002. Alteration of nuclear lamin organization inhibits RNA polymerase II-dependent transcription. *J Cell Biol*. 156:603-8.

Spector, D.L. 2003. The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem*. 72:573-608.

Spilianakis, C.G., M.D. Lalioti, T. Town, G.R. Lee, and R.A. Flavell. 2005. Interchromosomal associations between alternatively expressed loci. *Nature*. 435:637-45.

Sproul, D., N. Gilbert, and W.A. Bickmore. 2005. The role of chromatin structure in regulating the expression of clustered genes. *Nat Rev Genet*. 6:775-81.

Stastny, P. 2006. Introduction: MICA/MICB in innate immunity, adaptive immunity, autoimmunity, cancer, and in the immune response to transplants. *Hum Immunol*. 67:141-4.

Stephens, H.A. 2001. MICA and MICB genes: can the enigma of their polymorphism be resolved? *Trends Immunol*. 22:378-85.

Stern, L.J., J.H. Brown, T.S. Jardetzky, J.C. Gorga, R.G. Urban, J.L. Strominger, and D.C. Wiley. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*. 368:215-21.

Strick, R., Y. Zhang, N. Emmanuel, and P.L. Strissel. 2006. Common chromatin structures at breakpoint cluster regions may lead to chromosomal translocations found in chronic and acute leukemias. *Hum Genet*. 119:479-95.

Stuurman, N., S. Heins, and U. Aebi. 1998. Nuclear lamins: their structure, assembly, and interactions. *J Struct Biol*. 122:42-66.

Stuurman, N., A.M. Meijne, A.J. van der Pol, L. de Jong, R. van Driel, and J. van Renswoude. 1990. The nuclear matrix from cells of different origin. Evidence for a common set of matrix proteins. *J Biol Chem*. 265:5460-5.

Sullivan, T., D. Escalante-Alcalde, H. Bhatt, M. Anver, N. Bhat, K. Nagashima, C.L. Stewart, and B. Burke. 1999. Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J Cell Biol*. 147:913-20.

Sultmann, H., A. Sato, B.W. Murray, N. Takezaki, R. Geisler, G.J. Rauch, and J. Klein. 2000. Conservation of Mhc class III region synteny between zebrafish and human as determined by radiation hybrid mapping. *J Immunol*. 165:6984-93.

Sun, F.L., M.H. Cuaycong, and S.C. Elgin. 2001. Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. *Mol Cell Biol*. 21:2867-79.

Tang, J., P.N. Kao, and H.R. Herschman. 2000. Protein-arginine methyltransferase I, the predominant protein-arginine methyltransferase in cells, interacts with and is regulated by interleukin enhancer-binding factor 3. *J Biol Chem*. 275:19866-76.

Tatsumi, Y., S. Ohta, H. Kimura, T. Tsurimoto, and C. Obuse. 2003. The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. *J Biol Chem*. 278:41528-34.

Tenzen, T., T. Yamagata, T. Fukagawa, K. Sugaya, A. Ando, H. Inoko, T. Gojobori, A. Fujiyama, K. Okumura, and T. Ikemura. 1997. Precise switching of DNA replication timing in the GC content transition area in the human major histocompatibility complex. *Mol Cell Biol*. 17:4043-50.

Thompson, M., R.A. Haeusler, P.D. Good, and D.R. Engelke. 2003. Nucleolar clustering of dispersed tRNA genes. *Science*. 302:1399-401.

Thorsby, E. 1997. Invited anniversary review: HLA associated diseases. *Hum Immunol*. 53:1-11.

Tomazou, E.M., V.K. Rakyan, G. Lefebvre, R. Andrews, P. Ellis, C. Langford, L. Backdahl, M. Miretti, D. Ottaviani, D. Sheer, A. Murrell, and S. Beck. 2008. Generation of a strand-specific, genomic tiling array of the human Major Histocompatibility Complex (MHC) and its application for DNA methylation analysis. *BMC Medical Genomics*. 1:19.

Tremethick, D.J. 2007. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell*. 128:651-4.

Utley, R.T., K. Ikeda, P.A. Grant, J. Cote, D.J. Steger, A. Eberharter, S. John, and J.L. Workman. 1998. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature*. 394:498-502.

Vakoc, C.R., D.L. Letting, N. Gheldof, T. Sawado, M.A. Bender, M. Groudine, M.J. Weiss, J. Dekker, and G.A. Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell*. 17:453-62.

Valenzuela, L., and R.T. Kamakaka. 2006. Chromatin insulators. *Annu Rev Genet*. 40:107-38.

van den Elsen, P.J., S.J. Gobin, M.C. van Eggermond, and A. Peijnenburg. 1998. Regulation of MHC class I and II gene transcription: differences and similarities. *Immunogenetics*. 48:208-21.

van den Elsen, P.J., T.M. Holling, H.F. Kuipers, and N. van der Stoep. 2004. Transcriptional regulation of antigen presentation. *Curr Opin Immunol*. 16:67-75.

van Drunen, C.M., R.G. Sewalt, R.W. Oosterling, P.J. Weisbeek, S.C. Smeekens, and R. van Driel. 1999. A bipartite sequence element associated with matrix/scaffold attachment regions. *Nucleic Acids Res*. 27:2924-30.

Verschure, P.J., I. van der Kraan, E.M. Manders, D. Hoogstraten, A.B. Houtsmuller, and R. van Driel. 2003. Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO Rep*. 4:861-6.

Verschure, P.J., I. van Der Kraan, E.M. Manders, and R. van Driel. 1999. Spatial relationship between transcription sites and chromosome territories. *J Cell Biol*. 147:13-24.

Vishwanatha, J.K., H.K. Jindal, and R.G. Davis. 1992. The role of primer recognition proteins in DNA replication: association with nuclear matrix in HeLa cells. *J Cell Sci*. 101 (Pt 1):25-34.

Visser, A.E., F. Jaunin, S. Fakan, and J.A. Aten. 2000. High resolution analysis of interphase chromosome domains. *J Cell Sci*. 113 (Pt 14):2585-93.

Vogelstein, B., D.M. Pardoll, and D.S. Coffey. 1980. Supercoiled loops and eucaryotic DNA replicaton. *Cell*. 22:79-85.

Volpi, E.V., E. Chevret, T. Jones, R. Vatcheva, J. Williamson, S. Beck, R.D. Campbell, M. Goldsworthy, S.H. Powis, J. Ragoussis, J. Trowsdale, and D. Sheer. 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci*. 113 (Pt 9):1565-76.

von Kries, J.P., Buhrmester, H., Stratling, W., H. 1991. A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. *Cell*. 64:123-35.

von Kries, J.P., L. Phi-Van, S. Diekmann, and W.H. Stratling. 1990. A non-curved chicken lysozyme 5' matrix attachment site is 3' followed by a strongly curved DNA sequence. *Nucleic Acids Res*. 18:3881-5.

Wang, H., Z.Q. Huang, L. Xia, Q. Feng, H. Erdjument-Bromage, B.D. Strahl, S.D. Briggs, C.D. Allis, J. Wong, P. Tempst, and Y. Zhang. 2001. Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science*. 293:853-7.

Wang, J., C. Shiels, P. Sasieni, P.J. Wu, S.A. Islam, P.S. Freemont, and D. Sheer. 2004. Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *J Cell Biol*. 164:515-26.

Wang, J.C. 2002. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol*. 3:430-40.

Wang, T.Y. 1966. Solubilization and characterization of the residual proteins of the cell nucleus. *J Biol Chem*. 241:2913-7.

Ware, C.F. 2005. Network communications: lymphotoxins, LIGHT, and TNF. *Annu Rev Immunol*. 23:787-819.

Weighardt, F., F. Cobianchi, L. Cartegni, I. Chiodi, A. Villa, S. Riva, and G. Biamonti. 1999. A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock. *J Cell Sci*. 112 (Pt 10):1465-76.

Welcsh, P.L., and M.C. King. 2001. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet*. 10:705-13.

Wen, J., S. Huang, H. Rogers, L.A. Dickinson, T. Kohwi-Shigematsu, and C.T. Noguchi. 2005. SATB1 family protein expressed during early erythroid differentiation modifies globin gene expression. *Blood*. 105:3330-9.

Williams, R.R., S. Broad, D. Sheer, and J. Ragoussis. 2002. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp Cell Res*. 272:163-75.

Worman, H.J., and G. Bonne. 2007. "Laminopathies": a wide spectrum of human diseases. *Exp Cell Res*. 313:2121-33.

Woynarowski, J.M. 2004. AT islands - their nature and potential for anticancer strategies. *Curr Cancer Drug Targets*. 4:219-34.

Wright, K.L., and J.P. Ting. 2006. Epigenetic regulation of MHC-II and CIITA genes. *Trends Immunol*. 27:405-12.

Wu, J.R., and D.M. Gilbert. 1996. A distinct G1 step required to specify the Chinese hamster DHFR replication origin. *Science*. 271:1270-2.

Xu, M., and P.R. Cook. 2008. Similar active genes cluster in specialized transcription factories. *J Cell Biol*. 181:615-23.

Yamada, T., W. Fischle, T. Sugiyama, C.D. Allis, and S.I. Grewal. 2005. The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. *Mol Cell*. 20:173-85.

Yang, X.J., and E. Seto. 2007. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*. 26:5310-8.

Yusufzai, T.M., and G. Felsenfeld. 2004. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A*. 101:8620-4.

Zhang, Y., and D. Reinberg. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*. 15:2343-60.

Zhao, Z., G. Tavoosidana, M. Sjolinder, A. Gondor, P. Mariano, S. Wang, C. Kanduri, M. Lezcano, K.S. Sandhu, U. Singh, V. Pant, V. Tiwari, S. Kurukuti, and R. Ohlsson. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*. 38:1341-7.

Zhong, S., P. Salomoni, and P.P. Pandolfi. 2000. The transcriptional role of PML and the nuclear body. *Nat Cell Biol*. 2:E85-90.
Zika, E., and J.P. Ting. 2005. Epigenetic control of MHC-II: interplay between CIITA and histone-modifying enzymes. *Curr Opin Immunol*. 17:58-64.
Zippo, A., A. De Robertis, R. Serafini, and S. Oliviero. 2007. PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. *Nat Cell Biol*. 9:932-44.
Zou, Y., W. Bresnahan, R.T. Taylor, and P. Stastny. 2005. Effect of human cytomegalovirus on expression of MHC class I-related chains A. *J Immunol*. 174:3098-104.

Appendices

Appendix I - Position of clones constituted entirely of repetitive sequences

Start position	End position
29857045	29859021
29859685	29861731
29952327	29954750
29954618	29957153
30007795	30009507
30071694	30074112
30123120	30125633
30157712	30159479
30159325	30162190
30254297	30256821
30324946	30327580
30326343	30328780
30353431	30355824
30355557	30357404
30356186	30358959
30358915	30361227
30360627	30363086
30373125	30373916
30461074	30463679
30505856	30508413
30507610	30510119
30548127	30550788
30550719	30553560
30553123	30554913
30598279	30600699
30624293	30626447
31044408	31046885
31102983	31105539
31176621	31179067
31258475	31261419
31282010	31284107
31309154	31310921
31331246	31333604
31333527	31335887
31397092	31399917
31399278	31403895
31405420	31408180
31407745	31410625
31410312	31412594
31481894	31483829
31482771	31485215
31505209	31507825
31507238	31509555
31509279	31511644
31578226	31580534
31929098	31930771
32061007	32063565
32063457	32066073
32333324	32335035
32334902	32337196
32347172	32349492
32357072	32359919
32531595	32534000
32552075	32554386
32570284	32572093
32637066	32640334
32637942	32640486
32750587	32752380
32751956	32754798
32803977	32805579
32879848	32882714
32946350	32948854
33065138	33067726

33099436	33102017
33125489	33127433
33128270	33130623
33130078	33132394
33171878	33174286
33174199	33176819
33196915	33199698
33208179	33210083
33211780	33213487
33213814	33215866
33423446	33426238
33541330	33543791

Appendix II - Position of MARs of uninduced MRC5 cells and their array enrichment

Start position	End position	Matrix enrichment (log_2)
29802604	29804427	0.408556898
29811013	29813380	0.560802711
29840342	29842808	0.536847791
29849827	29852415	0.454672441
29872823	29875035	0.461270083
29882039	29884543	0.608527194
29884167	29886552	0.927214351
29894385	29896949	0.415990851
29933548	29935949	0.405509686
29937894	29939594	0.89732932
29960874	29962955	0.531348184
29975451	29978299	0.454830111
29982740	29984644	1.606179952
29996470	29998921	0.475552781
30036629	30039288	0.447158433
30046908	30048715	0.691269293
30062253	30064589	0.411414968
30065589	30067967	0.534589265
30101121	30103717	0.487064175
30103986	30106521	0.484163181
30129527	30132152	0.411419205
30131866	30134499	0.524926594
30134479	30136131	0.628146834
30146157	30148833	0.413190429
30168959	30171414	0.418168319
30200798	30203110	0.429058999
30223796	30226544	0.753265219
30226106	30228443	0.813896405
30231506	30234359	0.45203356
30238940	30240972	1.103889025
30249896	30252535	0.500941774
30252097	30254580	0.719450528
30304335	30306759	0.551212328
30341155	30343559	0.540194347
30343050	30345405	0.657699355
30377836	30380380	0.81314185
30389144	30392634	0.412102625
30391810	30393924	0.568857432
30421568	30424503	0.463906481
30428243	30430343	0.614378646
30442776	30445345	0.49071704
30455483	30457800	0.998706823
30459362	30461425	0.552074532
30472435	30474725	0.434562973
30482335	30484495	0.408831062
30501127	30503780	0.596996587
30507610	30510119	0.73442487
30509530	30512280	0.860153585
30516686	30519460	0.479511852
30610759	30613495	0.421833521
30698815	30701782	0.488345733
30731648	30734384	0.753858118
30767990	30770760	0.740066525
30772906	30775852	0.49751423
31056157	31059172	0.851972976
31092038	31094660	0.551624383
31321337	31323848	0.452281196
31323764	31326217	0.621960824
31365145	31367681	0.491564421
31367358	31369739	0.43577134
31389893	31392559	0.428661609
31415829	31418487	0.419018763
31425276	31427539	0.422257018
31511491	31514354	0.540576928
31865033	31867206	0.825494445
31955316	31957399	0.802678454

31975147	31977228	0.559284364
31980740	31983206	0.575024791
32220043	32222067	0.702016342
32364646	32367027	0.411144569
32387922	32390488	0.452297001
32390209	32393140	0.663247501
32395956	32398655	0.750176571
32409735	32412419	0.689602651
32425489	32427795	0.564261938
32430334	32433086	0.453128227
32434975	32437634	0.410066859
32449154	32450929	0.432421388
32471036	32473567	0.470043067
32498517	32501144	0.524899088
32505710	32507577	0.661459906
32509557	32512401	0.429390024
32511421	32513544	0.932602283
32526778	32528509	0.503215151
32539196	32541992	0.584627644
32561821	32564747	0.431994573
32566224	32568770	0.55458087
32583051	32585482	0.478776537
32619337	32622559	0.401899149
32622631	32625110	0.513908404
32651630	32653632	0.499003012
32686431	32689080	0.767205538
32691590	32694583	0.461439144
32705613	32708110	0.735658383
32761722	32764350	0.482341379
32769075	32771517	0.565704136
32780871	32783266	0.41115026
32787740	32790461	0.412398176
32794482	32797049	0.43454104
32796793	32799124	0.617260724
32857226	32859895	0.501158919
32870242	32872810	0.548254335
32875022	32877556	0.428879357
32877232	32879913	0.455350541
32904298	32906120	0.55790393
32931156	32933975	0.417799295
33025410	33027740	0.43507822
33027231	33028946	0.558000713
33067567	33070137	0.5299079
33104021	33106350	0.445969608
33117354	33120123	0.561697932
33176769	33178915	0.404808889
33217907	33219670	0.457952348
33227950	33230409	0.460856046
33283836	33286387	3.975215717
33292021	33293767	0.426293697
33501985	33503886	0.59069723

Appendix III - Position of MARs of IFN-γ induced MRC5 cells and their array enrichment

Start position	End position	Matrix enrichment (log_2)
29772399	29775033	1.049801067
29781700	29784102	0.591308658
29792469	29798300	0.522471833
29811013	29813380	0.420064147
29840342	29842808	0.641989126
29849827	29852415	0.704240122
29854767	29857057	0.552058507
29872823	29875035	0.54336622
29882039	29884543	0.834133308
29884167	29886552	1.133676991
29896602	29899084	0.441221525
29917144	29919923	0.50772475
29931599	29934065	0.529999226
29933548	29935949	0.910566307
29935654	29938224	0.74470164
29937894	29939594	0.441422054
29949751	29952328	0.463249917
29956443	29958865	0.617680963
29960874	29962955	1.158341109
29975451	29978299	0.50409617
29977965	29979873	0.425804085
29993954	29996553	0.644487391
29996470	29998921	0.58506104
30036629	30039288	0.442040222
30062253	30064589	0.541396708
30065589	30067967	0.844153915
30096934	30100100	0.447535682
30101121	30103717	0.639791775
30103986	30106521	0.665963702
30127345	30129803	0.493341502
30168959	30171414	0.516361315
30203030	30205314	0.50502874
30207839	30210162	0.604713047
30223796	30226544	0.762099579
30226106	30228443	1.771074308
30231506	30234359	0.722487581
30238940	30240972	2.140226418
30249896	30252535	0.641808161
30252097	30254580	0.68684891
30256578	30258948	0.418002505
30269344	30271071	0.461621459
30301929	30304620	1.081822769
30306687	30309185	0.415578249
30309105	30311651	0.574531723
30322904	30325267	0.43606955
30328752	30331413	0.550448473
30331401	30333872	0.606252271
30341155	30343559	0.655637831
30345571	30347925	0.545263234
30347804	30350169	0.46135727
30389144	30392634	0.594588791
30391810	30393924	0.754115027
30406289	30409007	1.121396851
30411277	30413752	0.542471092
30424262	30426665	0.513809249
30433512	30435982	0.469998703
30435458	30437915	0.559598092
30442776	30445345	0.458286542
30450571	30453097	0.515079905
30453257	30455963	1.092588766
30455483	30457800	1.266178864
30459362	30461425	0.757938884
30463598	30466352	0.45280234
30472435	30474725	0.590706527
30474644	30477030	0.488176961
30482335	30484495	0.522910204

30484481	30486798	0.495980716
30498694	30501279	0.582779522
30501127	30503780	0.956521698
30503512	30505932	0.52845754
30509530	30512280	0.818405019
30512150	30514491	0.595409879
30516686	30519460	0.583966874
30532105	30534885	0.433526001
30546495	30548204	0.800496996
30608279	30610773	0.426602843
30687802	30690256	0.717735973
30698815	30701782	0.441636431
30767990	30770760	0.925658748
30950261	30953164	0.403764536
31054143	31056020	0.421592821
31056157	31059172	0.554921445
31147190	31149630	0.412134893
31208146	31210054	0.587538815
31256049	31258433	0.512333365
31306864	31309399	0.449128931
31321337	31323848	0.83287629
31323764	31326217	0.769802675
31349918	31352712	0.425241291
31365145	31367681	1.001499443
31367358	31369739	1.052325164
31377505	31380155	0.549970869
31387943	31390679	0.90454688
31389893	31392559	1.020292464
31392265	31394713	0.546338622
31403562	31405377	0.446565181
31413230	31416000	0.914900848
31415829	31418487	0.668498238
31421730	31424152	0.414236451
31423462	31425292	0.493809738
31425276	31427539	0.418474252
31433998	31436656	0.545090692
31435990	31438961	0.436673775
31472505	31475307	0.455900346
31498765	31501407	0.401680246
31511491	31514354	0.77212005
31517732	31520447	0.742048175
31522428	31524251	0.589933091
31555714	31557747	0.415611255
31562288	31564901	0.407069889
31567390	31569981	0.403397752
31602298	31604502	0.617155371
31784795	31786680	0.40183233
31955316	31957399	0.494678105
31980740	31983206	0.613123821
31990195	31993322	0.531821297
32220043	32222067	0.727509267
32323976	32326426	0.638889147
32330237	32333661	0.405145132
32337055	32339933	0.65203743
32352121	32354558	0.40422516
32354408	32357209	0.432501891
32359755	32362494	0.443665253
32364646	32367027	0.611222219
32379696	32381542	0.401916558
32383418	32386054	0.406333698
32387922	32390488	0.6774036
32390209	32393140	0.428043767
32395956	32398655	1.597834078
32398586	32401154	0.891748454
32402691	32405190	0.447497692
32404753	32407328	0.458058934
32409735	32412419	0.864437415
32425489	32427795	0.772856453
32434975	32437634	0.419542132
32439560	32442112	0.523744186
32449154	32450929	0.73579992

32453538	32456020	0.519284458
32471036	32473567	0.647072857
32491122	32493607	0.447654425
32501095	32503289	0.454018427
32509557	32512401	0.587227372
32511421	32513544	2.50762754
32518443	32520782	0.544699667
32526778	32528509	0.664975997
32536608	32539324	0.616132169
32539196	32541992	1.027431171
32541661	32544149	0.401677674
32543851	32546046	0.461342381
32561821	32564747	0.793174255
32566224	32568770	0.478428081
32577343	32579881	0.430770385
32583051	32585482	0.493821162
32598134	32600400	0.412373568
32613182	32615215	0.456789805
32644074	32646662	1.030594313
32646521	32649544	0.501180242
32651630	32653632	0.448923754
32663263	32666061	0.404139716
32673701	32676631	0.414046656
32686431	32689080	0.866036275
32689045	32691474	1.184147147
32691590	32694583	0.559257989
32694444	32697209	0.46287668
32696991	32699506	0.547303174
32705613	32708110	0.738285211
32710133	32712269	0.423524638
32722734	32725314	0.495529479
32724504	32727177	0.490469236
32743165	32746083	0.404463712
32745804	32748726	0.430081816
32754319	32756817	0.580936359
32761722	32764350	0.754099217
32766998	32769591	0.615316648
32769075	32771517	0.719787252
32786130	32788393	0.686275865
32787740	32790461	0.927442836
32794482	32797049	0.50394299
32796793	32799124	0.62237831
32799001	32801669	1.4187691
32808098	32810522	0.598241428
32810462	32813136	0.420570067
32825193	32827718	0.586511952
32827724	32830188	0.515317133
32829829	32831757	0.461188079
32843585	32846054	0.418084091
32847268	32850579	0.692887483
32857226	32859895	0.778559184
32868718	32871367	0.539393651
32870242	32872810	0.479746347
32872192	32874816	0.437635
32875022	32877556	1.017812107
32877232	32879913	0.717756672
32931156	32933975	0.540953045
32941867	32944276	0.499969951
32962761	32965294	0.449659401
32974464	32976746	0.460148697
32975782	32978196	0.438711003
32977812	32980308	0.676242898
32980467	32982978	0.768576279
32992393	32994837	0.557045226
33022966	33025488	0.501840271
33025410	33027740	0.914369565
33067567	33070137	0.915400066
33090674	33093215	0.613554929
33104021	33106350	0.752821178
33115900	33118153	0.658419785
33117354	33120123	0.551879791

33121204	33123602	0.51283286
33123473	33125541	0.510379312
33133928	33138288	0.720228682
33138219	33141006	0.520126602
33142984	33145436	0.408756126
33176769	33178915	0.572146644
33192620	33193912	0.486828122
33210173	33212682	0.515207685
33227950	33230409	0.545032793
33320016	33322458	0.61226308
33462560	33465285	0.559465849
33501985	33503886	0.663283564
33529999	33532371	0.559354425

Appendix IV - Position of MARs of PGF cells and their array enrichment

Start position	End position	Matrix enrichment (log_2)
29739385	29741249	0.482147191
29755213	29757127	0.432167475
29849827	29852415	1.110277457
29861411	29863644	0.405538158
29894385	29896949	0.487401914
29896602	29899084	0.502859211
29929108	29931705	0.574111984
29937894	29939594	2.526421952
30023644	30025314	1.775267442
30053351	30055956	0.693329948
30065589	30067967	1.039362864
30101121	30103717	0.581704641
30103986	30106521	0.98209088
30106463	30109071	0.519614908
30108625	30111062	0.841034798
30110866	30113454	1.729229932
30113400	30115997	0.439612457
30118308	30120829	0.801492862
30123120	30125633	0.883641409
30125000	30127465	0.805619325
30127345	30129803	1.064832248
30129527	30132152	1.99585852
30131866	30134499	2.207236754
30134479	30136131	2.387396226
30152988	30155591	0.988586241
30155515	30157551	0.529730025
30200798	30203110	0.470841255
30219223	30221767	0.533182358
30231506	30234359	3.319040001
30249896	30252535	0.464204585
30252097	30254580	0.443791026
30314787	30317588	0.501078915
30341155	30343559	0.479846912
30345571	30347925	1.205872764
30347804	30350169	0.864868628
30350008	30351777	0.799927299
30363052	30365623	0.886622078
30365212	30368158	0.652169791
30368063	30370782	0.676448044
30368214	30370831	0.622433822
30377836	30380380	0.658218406
30380190	30382789	0.542408668
30391810	30393924	0.899687284
30448175	30450582	0.476275354
30463598	30466352	0.468579875
30474644	30477030	0.732587325
30482335	30484495	1.876325603
30489386	30491682	0.416035621
30498694	30501279	0.69955459
30501127	30503780	0.758140163
30503512	30505932	0.481793655
30509530	30512280	0.74238899
30555778	30557696	0.812807873
30557295	30559029	0.670945263
30562566	30564894	0.403426484
30568353	30570872	0.903176347
30583516	30586461	0.763093791
30600787	30603165	0.711956124
30610759	30613495	0.417564988
30677499	30679491	0.993747688
30689948	30692394	1.071166292
30691343	30693991	1.058426869
30698815	30701782	2.012782126
30701596	30704062	0.744320899
30711444	30713962	0.683121345
30714444	30716453	0.405939453
30756515	30758800	0.508400578
30767990	30770760	0.54222195
30772906	30775852	0.53441826

30825368	30827875	0.533282644
30827954	30830159	0.982094296
30834737	30837454	0.488754857
30901287	30903649	0.4370775
30903401	30906093	0.850043797
30923216	30925575	0.40814842
30991536	30994461	0.653581056
31018945	31021138	0.446628494
31021804	31024362	1.027783576
31023710	31026101	0.96905714
31024896	31027536	0.915876174
31026926	31029099	0.79052907
31028955	31031073	0.741360261
31030830	31033294	0.453431798
31039751	31042033	1.470046109
31049352	31051490	0.479301224
31051409	31054038	1.143856913
31054143	31056020	0.903290745
31056157	31059172	2.389286257
31061209	31064176	0.719780599
31098259	31100652	0.770924003
31100296	31102131	0.550215769
31105233	31107465	0.62682859
31110005	31112753	0.594295352
31174122	31176523	0.431525983
31256049	31258433	0.473060356
31323764	31326217	0.587113883
31352523	31355086	0.42361182
31369850	31372247	0.794692668
31373910	31376551	0.433074209
31376096	31377899	0.992805918
31403562	31405377	0.50653476
31413230	31416000	0.524884214
31435990	31438961	0.600006049
31438478	31440903	0.488539995
31442467	31444925	0.71193306
31444860	31447238	1.130968695
31451728	31454419	1.28228359
31454436	31456982	0.403756867
31485229	31487901	0.419517344
31490983	31493112	1.13447987
31511491	31514354	0.415125969
31522428	31524251	0.510902871
31558217	31560371	0.401852481
31575231	31578232	0.417827056
31588415	31591252	0.53639594
31621456	31623166	0.756629863
31662312	31665020	1.237122769
31676093	31678464	0.690785896
31678683	31681025	0.575932467
31687714	31690387	0.490032589
31689689	31692751	0.664720512
31796596	31799113	0.439280284
31838392	31840676	0.789680799
31938977	31941560	0.407787394
31945923	31948419	0.427803588
31955316	31957399	0.714972748
31963191	31965805	0.496633778
31980740	31983206	0.638547938
32193954	32196840	0.446889343
32220043	32222067	1.041134424
32246419	32248999	0.433639482
32248564	32250364	0.488978232
32249781	32252103	0.457409351
32323976	32326426	0.62996533
32327905	32330164	0.441934589
32339897	32342653	0.402688603
32359755	32362494	0.470886614
32372351	32375042	0.479955291
32376862	32379701	0.476129427
32386043	32387227	1.506569425
32393119	32395967	0.406674242

32411660	32414241	1.849664636
32414203	32416787	0.919980551
32416697	32418805	0.423529254
32418544	32420769	0.545012326
32423326	32425860	1.26078944
32432791	32435294	0.961228297
32434975	32437634	0.730141176
32439560	32442112	0.591598327
32446941	32449572	0.489568258
32453538	32456020	0.686435565
32467000	32470237	0.423336688
32483778	32485594	0.502526401
32496153	32498835	0.562715912
32498517	32501144	0.791099009
32505710	32507577	2.094044123
32524176	32526297	0.409230532
32526778	32528509	0.682785027
32536608	32539324	0.464144547
32543851	32546046	0.616745224
32566224	32568770	0.773844701
32577343	32579881	0.491508336
32577611	32579425	0.554686048
32583051	32585482	0.688380975
32585496	32588059	0.489706196
32587814	32590389	0.433742848
32613182	32615215	0.409983107
32619337	32622559	1.002131183
32626531	32628342	0.885650144
32629755	32631649	0.471231818
32649832	32652328	1.777332919
32653939	32656323	0.483995856
32657249	32658112	0.469473531
32668459	32673858	0.540204285
32678591	32681243	0.662968739
32680961	32683760	0.553234079
32710133	32712269	0.670335726
32714274	32717018	1.377067649
32716744	32719125	0.693679936
32718336	32721027	0.463576216
32720216	32722920	0.836576859
32722734	32725314	0.622251611
32724504	32727177	0.452963123
32743165	32746083	0.524197086
32759416	32762298	2.056098071
32761722	32764350	0.431556292
32766998	32769591	0.645115482
32769075	32771517	1.12835019
32771813	32774032	3.031142095
32773440	32776301	0.499332172
32776173	32778703	3.341811971
32778597	32781168	1.112091941
32786130	32788393	1.279673698
32787740	32790461	1.534303387
32794482	32797049	1.04283508
32796793	32799124	0.505040719
32801443	32804134	0.597150639
32805771	32808229	1.192730415
32808098	32810522	0.793411988
32810462	32813136	0.52587137
32812745	32815129	0.531638095
32827724	32830188	0.640426903
32843585	32846054	0.409749519
32847268	32850579	0.608774574
32850451	32852327	1.587074053
32857226	32859895	0.609320612
32859507	32861730	2.525040351
32861678	32864080	1.060303875
32877232	32879913	0.540065293
32890054	32892994	0.556003338

Appendix V - Position of MARs common to all cell types

Start position	End position
29849827	29852415
29937894	29939594
30065589	30067967
30101121	30103717
30103986	30106521
30231506	30234359
30249896	30252535
30341155	30343559
30391810	30393924
30482335	30484495
30501127	30503780
30509530	30512280
30698815	30701782
30767990	30770760
31056157	31059172
31323764	31326217
31511491	31514354
31955316	31957399
31980740	31983206
32220043	32222067
32434975	32437634
32526778	32528509
32566224	32568770
32583051	32585482
32761722	32764350
32769075	32771517
32787740	32790461
32794482	32797049
32796793	32799124
32857226	32859895
32877232	32879913
33117354	33120123
33501985	33503886

Appendix VI - Position of MARs in the MHC of both IFN-γ induced and uninduced MRC5 cells

Start position	End position
29811013	29813380
29840342	29842808
29872823	29875035
29882039	29884543
29884167	29886552
29933548	29935949
29960874	29962955
29975451	29978299
29996470	29998921
30036629	30039288
30062253	30064589
30168959	30171414
30223796	30226544
30226106	30228443
30238940	30240972
30389144	30392634
30442776	30445345
30455483	30457800
30459362	30461425
30472435	30474725
30516686	30519460
31321337	31323848
31365145	31367681
31367358	31369739
31389893	31392559
31415829	31418487
31425276	31427539
32364646	32367027
32387922	32390488
32390209	32393140
32395956	32398655
32409735	32412419
32425489	32427795
32449154	32450929
32471036	32473567
32509557	32512401
32511421	32513544
32539196	32541992
32561821	32564747
32651630	32653632
32395956	32398655
32686431	32689080
32691590	32694583
32705613	32708110
32875022	32877556
32931156	32933975
33025410	33027740
33067567	33070137
33104021	33106350
33176769	33178915
33227950	33230409

Appendix VII - Position of MARs of PGF cells only

Start position	End position
29739385	29741249
29755213	29757127
29861411	29863644
29929108	29931705
30023644	30025314
30053351	30055956
30106463	30109071
30108625	30111062
30110866	30113454
30113400	30115997
30118308	30120829
30125000	30127465
30152988	30155591
30155515	30157551
30219223	30221767
30314787	30317588
30350008	30351777
30363052	30365623
30365212	30368158
30368063	30370782
30368214	30370831
30380190	30382789
30448175	30450582
30489386	30491682
30555778	30557696
30557295	30559029
30562566	30564894
30583516	30586461
30600787	30603165
30677499	30679491
30689948	30692394
30691343	30693991
30701596	30704062
30711444	30713962
30714444	30716453
30756515	30758800
30825368	30827875
30827954	30830159
30834737	30837454
30901287	30903649
30903401	30906093
30923216	30925575
30991536	30994461
31018945	31021138
31021804	31024362
31023710	31026101
31024896	31027536
31026926	31029099
31028955	31031073
31030830	31033294
31039751	31042033
31049352	31051490
31051409	31054038
31054143	31056020
31061209	31064176
31098259	31100652
31100296	31102131
31105233	31107465
31110005	31112753
31174122	31176523
31352523	31355086
31369850	31372247
31373910	31376551
31376096	31377899
31438478	31440903

31442467	31444925
31444860	31447238
31451728	31454419
31454436	31456982
31485229	31487901
31490983	31493112
31558217	31560371
31575231	31578232
31588415	31591252
31621456	31623166
31662312	31665020
31676093	31678464
31678683	31681025
31687714	31690387
31689689	31692751
31796596	31799113
31838392	31840676
31938977	31941560
31945923	31948419
31963191	31965805
32193954	32196840
32246419	32248999
32248564	32250364
32249781	32252103
32327905	32330164
32339897	32342653
32372351	32375042
32376862	32379701
32386043	32387227
32393119	32395967
32411660	32414241
32414203	32416787
32416697	32418805
32418544	32420769
32423326	32425860
32432791	32435294
32446941	32449572
32467000	32470237
32483778	32485594
32496153	32498835
32524176	32526297
32577611	32579425
32585496	32588059
32587814	32590389
32626531	32628342
32629755	32631649
32653939	32656323
32657249	32658112
32668459	32673858
32678591	32681243
32680961	32683760
32714274	32717018
32716744	32719125
32718336	32721027
32720216	32722920
32759416	32762298
32771813	32774032
32773440	32776301
32776173	32778703
32778597	32781168
32801443	32804134
32805771	32808229
32812745	32815129
32850451	32852327
32859507	32861730
32861678	32864080
32890054	32892994
32892523	32894937

32894483	32897025
32896974	32899399
32906204	32908888
32944237	32946032
32948866	32951226
32951208	32953571
32964422	32967273
32967276	32969811
32968550	32971099
32970390	32972719
32972559	32975260
32987255	32990019
32989946	32992530
32997141	32999893
32999284	33001846
33003704	33006056
33006053	33008988
33008444	33010878
33010869	33013520
33034370	33036962
33037346	33039646
33039760	33042545
33049800	33052268
33086287	33088714
33092818	33095473
33097613	33099941
33106390	33108236
33109482	33111869
33112049	33114719
33113667	33116173
33144119	33146299
33146226	33148958
33148418	33151260
33150708	33153101
33162178	33164006
33219995	33222707
33538830	33541653

Appendix VIII - Position of MARs of uninduced MRC5 cells

Start position	End position
29802604	29804427
29894385	29896949
29982740	29984644
30046908	30048715
30129527	30132152
30131866	30134499
30134479	30136131
30146157	30148833
30200798	30203110
30304335	30306759
30343050	30345405
30377836	30380380
30421568	30424503
30428243	30430343
30610759	30613495
30731648	30734384
30772906	30775852
31092038	31094660
31865033	31867206
31975147	31977228
32430334	32433086
32498517	32501144
32505710	32507577
32619337	32622559
32622631	32625110
32780871	32783266
32904298	32906120
33027231	33028946
33217907	33219670
33283836	33286387
33292021	33293767

Appendix IX - Position of MARs of IFN-γ induced MRC5 only

Start position	End position
29772399	29775033
29781700	29784102
29792469	29798300
29854767	29857057
29917144	29919923
29931599	29934065
29935654	29938224
29949751	29952328
29956443	29958865
29977965	29979873
29993954	29996553
30096934	30100100
30203030	30205314
30207839	30210162
30256578	30258948
30269344	30271071
30306687	30309185
30309105	30311651
30322904	30325267
30328752	30331413
30331401	30333872
30406289	30409007
30411277	30413752
30424262	30426665
30433512	30435982
30435458	30437915
30450571	30453097
30453257	30455963
30484481	30486798
30512150	30514491
30532105	30534885
30546495	30548204
30608279	30610773
30687802	30690256
30950261	30953164
31147190	31149630
31208146	31210054
31306864	31309399
31349918	31352712
31377505	31380155
31387943	31390679
31392265	31394713
31421730	31424152
31423462	31425292
31433998	31436656
31472505	31475307
31498765	31501407
31517732	31520447
31562288	31564901
31567390	31569981
31602298	31604502
31784795	31786680
31990195	31993322
32330237	32333661
32337055	32339933
32352121	32354558
32354408	32357209
32379696	32381542
32383418	32386054
32398586	32401154
32402691	32405190
32404753	32407328

32491122	32493607
32501095	32503289
32518443	32520782
32541661	32544149
32598134	32600400
32644074	32646662
32646521	32649544
32663263	32666061
32673701	32676631
32689045	32691474
32694444	32697209
32696991	32699506
32745804	32748726
32754319	32756817
32799001	32801669
32825193	32827718
32829829	32831757
32868718	32871367
32872192	32874816
32962761	32965294
32974464	32976746
32975782	32978196
32980467	32982978
33022966	33025488
33090674	33093215
33115900	33118153
33121204	33123602
33123473	33125541
33142984	33145436
33192620	33193912
33210173	33212682
33320016	33322458
33462560	33465285
33529999	33532371

Appendix X - Position of MARs of both IFN-γ induced MRC5 and PGF cells

Start position	End position
29896602	29899084
30127345	30129803
30345571	30347925
30347804	30350169
30463598	30466352
30474644	30477030
30498694	30501279
30503512	30505932
31054143	31056020
31256049	31258433
31403562	31405377
31413230	31416000
31435990	31438961
31522428	31524251
32323976	32326426
32359755	32362494
32439560	32442112
32453538	32456020
32536608	32539324
32543851	32546046
32577343	32579881
32613182	32615215
32710133	32712269
32722734	32725314
32724504	32727177
32743165	32746083
32766998	32769591
32786130	32788393
32808098	32810522
32810462	32813136
32827724	32830188
32843585	32846054
32847268	32850579
32941867	32944276
32977812	32980308
32992393	32994837
33133928	33138288
33138219	33141006